Ph. D. thesis

"Genome engineering in *Escherichia coli*: methods and applications"

Vitaliy S. Kolisnychenko

Ph. D. supervisor Dr. Pósfai György

Institute of Biochemistry
Biological Research Center
Hungarian Academy of Sciences

Szeged
2002

Contents

Acknowledgements

I did not compose, and certainly could not have composed, this thesis without generous contributions from many people.

First of all, I would like to thank my Ph. D. advisor Dr. Pósfai György for his scientific guidance and his support throughout my stay here, in Hungary.

I am very grateful to all past and present members of the laboratory, Zsuzsa Bereczki, Umenhoffer Kinga, Fehér Tamás, for their help and friendship. I would like to express special thanks to Szalkanovics Agnes for the excellent technical assistance.

I would like to thank Dr. Udvardy Andor and Dr. Duda Ernő for their valuable comments and suggestions about my thesis.

I would like to say thank you to all my friends in BRC in their continuous support and friendship.

I would like to thank our collaborators in the United States of America.
Dr. Frederick R. Blattner, who provided financial and intellectual support of current work.
Dr. Guy Plunkett III, who performed genomic analysis and identification of the segments to be deleted and Christopher D. Herring for helpful discussions.
The scarless deletion method was designed by Dr. Pósfai György, who also took part in some complex experiments.

Abbreviations

wt- wild type

PCR- polymerase chain reaction

Taq- *Thermus aquaticus*

bp- basepairs

ss- single-stranded

ds- double-stranded

DSB- double-stranded break

kb- kilobasepairs

Mb- megabasepairs

nt- nucleotide(s)

oligo- oligodeoxyribonucleotides

IS- insertion sequence

CSM- counter selection marker

ts- temperature sensitive

Ab- antibiotic

1. Introduction

Escherichia coli is a prokaryotic organism well known as both a tool and object for genome science studies. Genetic and biochemical data of the *E. coli* cell have been accumulating for over fifty years. Furthermore, as the primary model organism for bacteria it was used to elucidate the biological processes responsible for basic cellular functions. Gene annotations of this bacterium are regularly transferred to other organisms as their genome sequences are completed. The *E. coli* cell is also used extensively as a genomics tool to propagate DNA subclones for sequencing and for a variety of functional studies in many species. Industrially, *Escherichia coli* is used to produce hormones, enzymes and antibiotics. However, in spite of the vast knowledge of this bacterium, much remains to be learned about the composition of its genome.

The genome sequence of *E. coli* K-12 has been published in 1997 (Blattner et al. 1997), followed by the sequence of the pathogenic O157:H7 in 2001 (Perna et al. 2001). The availability of the genome sequences, together with the extensive genetic and biochemical data open the way for the targeted engineering of the *E. coli* genome. The thesis describes two novel genome engineering methods. These methods can serve multiple goals in functional genomics studies. On one hand, single genes can be replaced by mutant alleles in order to study their function. On the other hand, genes or entire genomic regions can be deleted from the chromosome to construct an improved model organism and biotechnological tool.

1.1 The *Escherichia coli* genome

In nature, *E. coli* is a facultative anaerobe and a metabolic opportunist spending part of its natural life cycle living anerobically in the intestinal tracts of animals and part living aerobically in diverse natural environments, such as rivers, soil or in the case of pathogens invading animal or human hosts (Schaechter and Neidhardt 1987). Individual strains of *E. coli* vary in their preferences for niches/hosts, and these variations are reflected in their gene contents. It has been estimated that the genomes of natural isolates of *E. coli* range from 4.5 to 5.5 megabases (Bergthorsson and Ochman 1998).

Two strains of *E. coli* have been fully sequenced, the K-12 strain MG1655 with 4,638,858 basepairs (Blattner et al. 1997) and two isolates of the enteric pathogen O157:H7 of 5,528,445 basepairs (Hayashi et al. 2001; Perna et al. 2001). K-12 and O157:H7 are phylogenetically distant relatives within the *E. coli* species (Reid et al. 2000, Perna et al. 2001).

The genome of the laboratory strain K-12 MG1655 is composed of nearly 4300 genes. Based on experimental data and homology searches, the majority of them can be identified. However, over one-third of the genes have no function assigned to. Protein coding sequences account for 88% of the genome, 0,8% encodes stable RNAs, 0,7% consists of non-coding repeats, leaving 11% for regulatory and other functions.

Comparison of the genomes of K-12 and O157:H7 revealed a startling pattern in which hundreds of strain-specific "islands" are found inserted into a common "backbone" that is highly conserved (98%) between the two strains. The strain specific islands are termed K-islands or O-islands after the strain from which they were sequenced. The size of the "backbone" genome is about 3.7 megabasepairs and the total of K-islands amounts to 0.9 Mbp. Thus removal of all the K-islands from MG1655 would result in a 20% reduction of the length of the genome.

Gene loss and horizontal gene transfer were the major genetic processes that shaped the ancestral *E. coli* genome resulting in the spectrum of divergent present-day strains possessing very different arrays of genes (McClelland et al. 2000, Ochman and Jones 2000, Riley and Serres 2000, Perna et al. 2001). The genes contained in the backbone regions generally include basic core functions of *E. coli* that are necessary regardless of environmental niche. Islands contain a disproportionate share of genes that are of unknown function as well as toxins, virulence factors and metabolic capabilities that may be of advantage in the niche to which the strain is adapted. Islands also contain many transposable elements, phages, cryptic prophages, pseudogenes, gene remnants, and damaged operons.

Clearly *E. coli* possesses many dispensable functions that would not be needed in a strain designed for laboratory purposes. Comparative genomics provides the working hypothesis that genes and gene islands that are present in MG1655 and not in other strains may be deleted. A second approach is to eliminate genes that are not expressed in

any of a wide variety of conditions. Based on microarray data, about 25% of *E. coli* genes are clearly expressed in log phase growth in minimal medium and 50% appear not to be with the remainder probably expressed at a low level. (Richmond et al. 1999, and data not shown).

A minimal strain consisting just of the backbone, about 3700 genes, would be interesting to study. Many questions of bacterial physiology, genetics and ecology could be addressed experimentally by using a simplified "clean" *E. coli* strain. Would such a strain be healthier than its parent MG1655? Would it have a more stable genome? What horizontal transmissions would it pick up from the environment? Does "selfish DNA" extract a cost?

Potential advantages of a reduced genome in biotechnology are also readily apparent. Any unnecessary gene product that is expressed in a production host represents a potential contaminant that could drive up the cost of product purification. There is also a metabolic waste entailed in producing unwanted products. Some side products are detrimental even in tiny quantities when drugs or vaccines must pass certification. Deletion of the gene is by far the most reliable and effective way to ensure the complete absence of an unwanted component in a biotechnology product.

Our laboratory is particularly interested in aspects of evolution and stability of the genome. Major changes in the composition and structure of the genome are mediated by mobile genetic elements. *E coli* MG1655 includes 44 transposable elements of 10 kinds and 6 copies of Rhs (Bachellier et al. 1996), a repeated sequence which resembles a transposon but for which there is no data indicating that it can move. Transposases associated with the transposable elements are induced by stress such as heat and cold shock and might be activated by procedures such as electroporation. Genome rearrangements and mutations associated with the activity of transposons are very common in *E. coli* including the preponderance of spontaneous null mutations (Kitamura et al. 1995) as well as inversions. The genome also contains 8 prophage or phage remnants (Campbell et al. 1996, Retallack et al. 1994). These are thought to be the primary vehicles of horizontal gene-transfer events.

The long-term goal of the laboratory is the construction of a reduced, backbone-only *E. coli* genome in order to study the role of mobile genetic elements in intra- and

intergenomic rearrangements, and the effects of such events in adaptation and evolution. Construction of a reduced genome requires new genome engineering methods, which are rapid, efficient, permit the introduction of multiple deletions, and leave no extra sequences ("scars") behind.

1.2 Genome engineering methods

The availability of the genome sequence permits the introduction of practically any targeted modification (insertion, deletion, point mutation) into the *E. coli* chromosome. A variety of genetic tools are used by different laboratories in experiments requiring genomic modifications. In most cases, the general scheme consists of the following steps: (i) engineering of a targeting DNA-construct (plasmid or linear fragment) (ii) insertion of the DNA-construct into a selected locus of the genome by homologous recombination (iii) removal of a segment of the genome overlapping the inserted sequence. The two basic variations of this scheme, relevant to the subject of the thesis, are reviewed below.

1.2.1 RecA-dependent suicide plasmid-based genome engineering methods

Outline of the method

The general scheme of the RecA-dependent suicide plasmid-based genome engineering method is depicted on Figure 1. First, a mutant allele is constructed by in vitro recombinant DNA techniques. This fragment, typically 500-1000 bp long, is then cloned into a suicide plasmid vector (Hamilton et al. 1989, Blomfield et al. 1991, Link et al. 1997), and the plasmid construct is transformed into the target cell. Under non-permissive conditions, the plasmid is unable to replicate, and is quickly eliminated from the cell. However, in a fraction of the cells, the plasmid integrates into the chromosome via homologous recombination between the wild-type allele on the chromosome and the mutant allele on the plasmid (O'Connor et al. 1989, Poustka et al. 1984). This recombination event is mediated by RecA (Clark et al. 1965, Howard-Flanders et al. 1966), the enzyme involved in the recombination and repair processes of the cell. Cells, which carry the plasmid integrated into the genome, can be isolated by proper antibiotic selection. The cointegrate structure of these cells is characterized by a duplicated

sequence (mutant and wt alleles) flanking the plasmid segment. Spontaneous resolution of the cointegrate by homologous recombination between the duplicated sequences can result either in recreation of the wt allele, or conversion of the locus into the mutant allele (Hamilton et al. 1989, Blomfield et al. 1991, Link et al. 1997). Since spontaneous resolution of the cointegrate is a rare event, efficient counterselection against a plasmid-coded feature is needed to eliminate all cells retaining the plasmid sequences in the genome. As a final step, using proper screening methods (e.g., allele-specific PCR), the two outcomes of the resolution process can be distinguished, and cells carrying the mutant allele can be isolated.

Bacterial allelic exchange using recombinant suicide vectors has been extensively exploited to introduce recombinant or mutated alleles into the chromosomes of both gram-positive and gram-negative bacteria (Maloy et al. 1996, Snyder et al. 1996).

Suicide plasmids

Suicide plasmids are characterized by their conditional replication (Hamilton et al. 1989). They can multiple in the cloning host, but their replication in the target cell depends on external factors. The most widely used suicide plasmids utilize a temperature-sensitive mutant of the pSC101 replicon (Frey et al. 1979, Hasunuma et al. 1977, 1979). These plasmids can multiple at 30°C, but their replication ceases at 37-42°C. Another set of suicide plasmids are based on the R6K origin. Replication of these plasmids requires the presence of the Π protein (Filutowicz et al. 1986, Germino et. al, 1983, Kelley et al. 1985). Expression of Π can be controlled by applying an inducible promoter. The suicide plasmids used in this study are based on the plasmid set developed in this laboratory (Pósfai et al. 1994). These plasmids are characterized by a modular composition, and offer a choice of three antibiotic resistance genes (Ap, Km, Cm) and two conditional replication system (pSC101 ts, R6K).

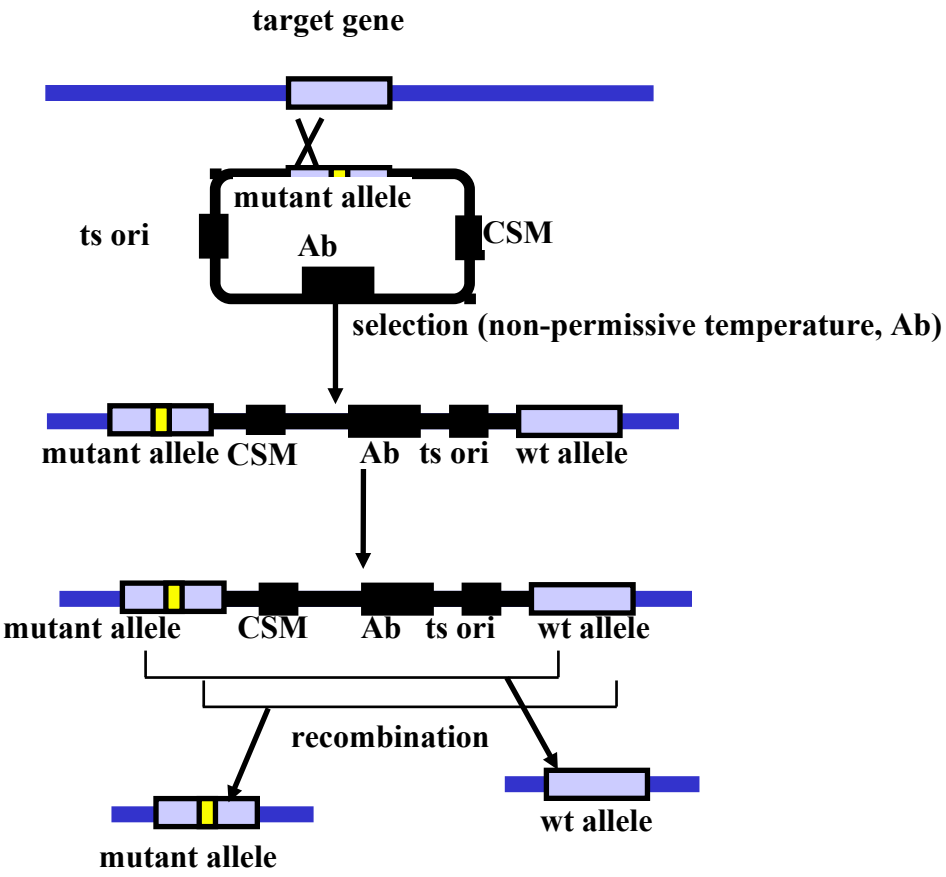


Figure 1. General strategy of RecA-dependent suicide plasmid-based genome engineering. Cointegrates of the chromosome and the suicide plasmid can be formed via homologous recombination between the mutant and the wt alleles of the target gene. The plasmid carries a temperature-sensitive (ts) replicon, an antibiotic (Ab) resistance gene and counter selection marker (CSM). Cointegrates are selected by their antibiotic resistance at the non-permissive temperature for plasmid replication. Spontaneous recombination between the duplications results either in a reversion to the wt chromosome or in a markerless gene replacement event.

Counterselection methods

The presence of the counterselection marker gene in the plasmid, integrated into the chromosome, serves as negative selection marker (Dean et al. 1981, Gay et al. 1985,

Russell et al. 1989). The most widely used genes to promote the selection for mutated alleles are *rpsL* (Skorupski et al. 1996, Stibitz et al. 1986), *tetA* (Maloy et al. 1981), and *sacB* (Gay et al. 1985, Ried et al. 1987), causing sensitivity to streptomycin, fusaric acid and sucrose, respectively.

Advantages and difficulties of the method

The long, homologous targeting fragment of the suicide plasmid is recombined into the genome by RecA with relatively high efficiency. Even so, a high number of cells ($>10^6$) must be transformed with the suicide plasmid in order to obtain a recombinant. To overcome this difficulty, suicide plasmids with the pSC101 ts replicon are first maintained and replicated at permissive temperatures in the target cell, then the replication is turned off by elevating the temperature to create non-permissive conditions.

The procedure does not produce a mutant directly. Rather, the result is a mixture of cells carrying wt or mutant alleles. Efficient screening systems must be used to isolate a desired mutant.

The procedure is relatively laborious and slow, due to the cloning steps required for ligating a targeting DNA-fragment into the suicide plasmid.

The biggest disadvantage of the method is the need for a counterselection procedure. In our hands, even the most widely used system, sacB/sucrose counterselection, produced a high non-specific background of cells. All counterselection methods suffer from medium-, temperature- and strain-dependence (Maloy et al. 1996, Miller et al. 1988, Blomfield et al. 1991, Link et al. 1997).

Double-strand break stimulated gene replacement method

A new variation of the RecA-dependent, suicide plasmid-based method has recently been developed in our laboratory. The novelty of the method lies in the use of meganuclease I-*Sce*I. The meganuclease causes a double-strand break (DSB) in the DNA at its recognition site. Since the recognition site is extremely large (18 bp), it does not exist in most genomes, including *E. coli* K-12. In the new method, the suicide plasmids were engineered to carry a I-*Sce*I site. After integration of the suicide plasmid into the genome, I-*Sce*I expression is induced from a helper plasmid. The enzyme cleaves its site in the plasmid sequence, causing a DSB in the chromosome. Free DNA ends stimulate recombinational repair of the gap, and this repair is most likely to occur between the

homologous segments provided by the wt and mutant allele pair. I-*Sce*I cleavage thus serves two purposes: (i) stimulates resolution of the cointegrate by providing free DNA ends for recombination (ii) acts as a general counterselection tool (not limited by medium-, temperature- and strain-specificity) by eliminating cells having a genome with the inserted cleavage site.

1.2.2 λ Red-dependent genome engineering methods

Outline of the method

Alternative recombination systems, such as λ Red, have recently been used to create gene replacements, genomic insertions or deletions. First, a linear DNA-fragment is generated by PCR on an appropriate template plasmid. The fragment carries a selectable marker gene and short (40-50 bp) terminal sequences which are homologous to arbitrarily chosen genomic regions. These terminal homologous segments are incorporated into the fragment as 5' parts of the synthetic PCR primers. The linear DNA-fragment is then electroporated into the target cell where the fragment can replace the selected genomic region by a double crossover mediated by the terminal homologous segments (Jasin et al. 1984, Marinus et al. 1983, Russell et al. 1989, Winans et al. 1985). To achieve detectable level of recombination, the cell must express the λ *red αβγ* genes (or similar), whose products are called Exo, β and γ, respectively (Murphy, 1998). γ prevents degradation of the linear fragment by inhibiting the host RecBCD exonuclease V (Lorenz et al. 1994), so that β and Exo can gain access to the DNA-fragment ends to promote recombination.

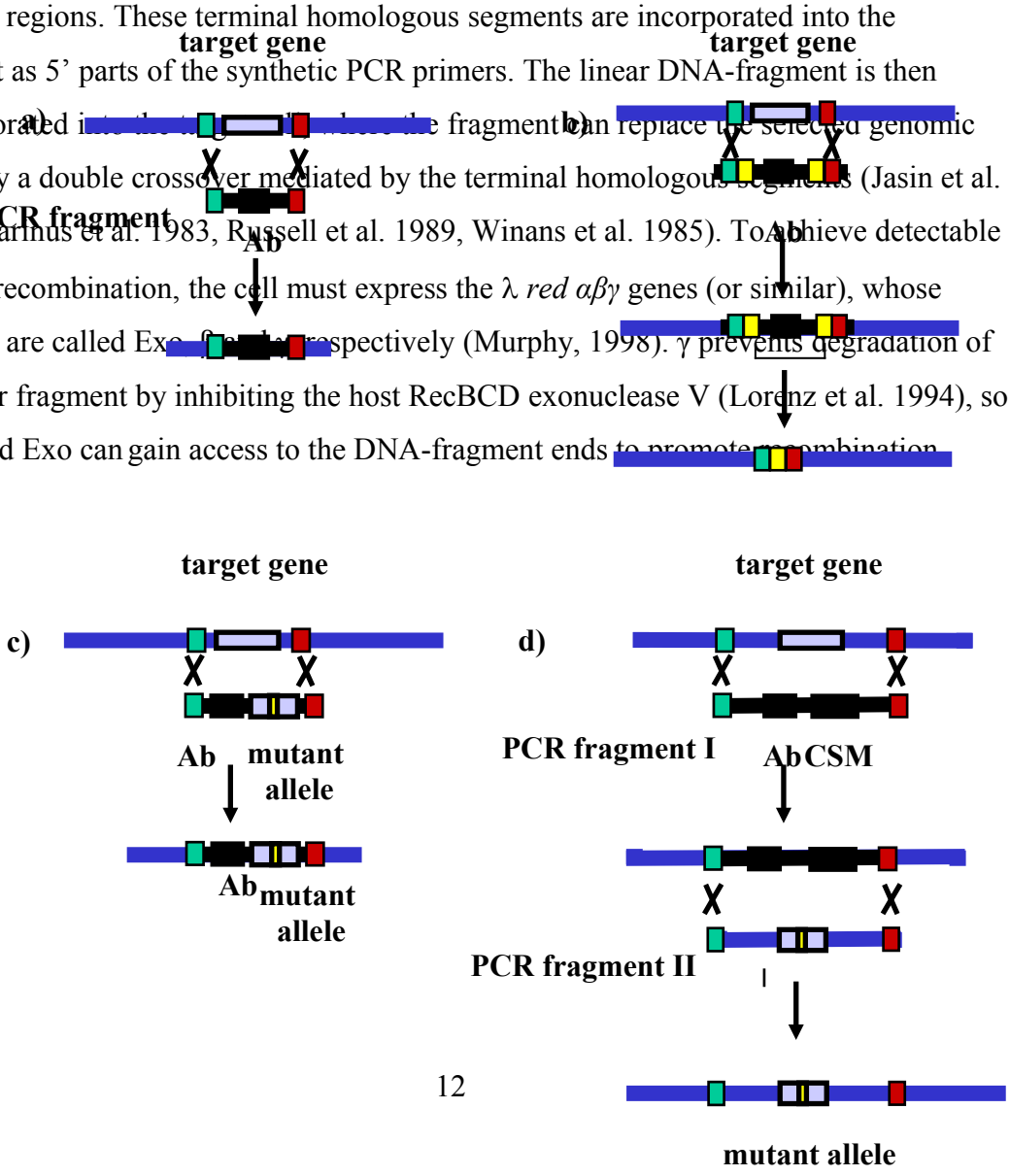


Figure 2. Four generally used schemes for generation of knock-out genes (a, b) or gene replacements (c, d) in *E. coli*. a) Simple gene knockout using a PCR fragment containing an antibiotic resistance marker. b) Gene knockout using a PCR fragment, followed by excision of the antibiotic resistance marker via specific recombination between *frt* or *loxP* sites (yellow boxes), mediated by Flp/Cre recombinase. c) Gene replacement using a PCR fragment, which carries an antibiotic resistance gene and a mutant allele of the target gene; d) Two-round markerless gene replacement. Red and green boxes represent two different homologous segments.

Advantages and difficulties of the method

The method is fast, because there is no need for cloning steps, and the DNA-fragment can be generated by a simple PCR.

In contrast to the suicide plasmid-based method, the method results directly in a mutant construct, eliminating the need for a screen to distinguish wt and mutant cells.

A potential drawback is that a helper plasmid (or specifically engineered chromosome) must carry the λ *red* genes. Preferentially, these genes are expressed in an inducible manner, and are turned on only when needed. Constitutively expressed λ *red* genes can promote unwanted genomic rearrangements between small repeat sequences.

A simple replacement of a genomic region results in the insertion of foreign sequences (e.g., selection marker gene) into the chromosome (Figure 2.). Removal of these sequences requires further manipulations. The selection marker gene, if flanked by proper recombinase recognition sites (*frt*, *loxP*), can be excised by inducing a specific recombinase (Flp, Cre). However, a single *frt* or *loxP* site will remain on the chromosome, potentially preventing multiple rounds of consecutive gene replacements or deletions. Alternatively, a second linear fragment engineered to carry the desired "clean" genomic sequence can be used to replace the first one in the genome, but this procedure is laborious and cost-inefficient. In addition, it requires the use of a counterselection marker, posing problems of medium-, temperature- or strain-dependence (Cosloy et al. 1973, Clark et al. 1994, Winans et al. 1985, Amundsen et al. 1986, Biek et al. 1986, Russell et al. 1989).

2. Objectives

The primary goal of the laboratory is to engineer a simplified *E. coli* genome to study factors of genome evolution and stability. Specific goals relevant to this thesis:

1. Explore the use of the DSB-stimulated gene replacement method in producing genomic deletions.
2. Develop a rapid, efficient and "scarless" (no external sequences are left in the genome) deletion method based on the use of PCR-generated linear DNA-fragments combined with I-*Sce*I-stimulated DSB repair.
3. Study various factors, which influence the efficiency of the scarless deletion method; find the optimal conditions.
4. Introduce a series of consecutive deletions into the genome of K-12 MG1655 in order to create a backbone-only genome. Deletion targets should be identified by comparative genomics. Primary deletion candidates include agents of horizontal gene transfer and other genomic rearrangements (prophages, phage remnants, insertion sequences [ISs]) and major K-12-specific islands.

3. Materials and methods

3.1 Bacterial strains and media

Plasmids were generally prepared from DH5α (Woodcock et al. 1989). Deletions were engineered in the strain K-12 MG1655 (Blattner et al. 1997). Standard laboratory media (LB, SOC, and M9/glucose minimal medium) and agar plates were used (Sambrook et al. 1989). Casamino acids were added to 0.5% when used as a medium supplement. Antibiotics were applied at the following concentrations: ampicillin (Ap) 50 μg/mL, chloramphenicol (Cm) 25 μg/mL, and kanamycin (Km) 25 μg/mL. Heat-treated chlor-tetracycline (cTc) was used to inactivate the Tet repressor. Preparation of the inducer stock solution has been described (Pósfai et al. 1999).

3.2 Plasmids

3.2.1 Plasmids used in the DSB-stimulated deletion method

The I-SceI site-containing suicide plasmid series pST76/pSG76 has been developed in this laboratory and has been described (Pósfai et al. 1997, 1999). pST76-K and pSG76-C, used in this study, are available under accession numbers Y09897 and Y09893, respectively. Plasmid pST76-K carries a temperature-sensitive replicon and cannot replicate at 37-42°C (Pósfai et al. 1997, Armstrong et al. 1984). Plasmid pSG76-C requires the product of the *pir* gene - Π protein - for replication (Pósfai et al. 1997, Stalker et al. 1982). Π protein is expressed from helper plasmid pFT-A (Pósfai et al. 1997).

To construct targeting suicide plasmid pSG76-CMB45, a 542-bp *Bgl*II-*Cla*I fragment of the PCR product obtained by primers M1/M7 on MG1655 genomic DNA template was cloned between the *Bam*HI and *Cla*I site of pSG76-C, followed by ligation of a 614-bp *Cla*I fragment of another PCR product, obtained with primers M4/M5, into the *Cla*I site. This latter *Cla*I fragment was cloned into the *Cla*I site of pST76-K to obtain a second targeting plasmid, pST76-KM45.

The I-*Sce*I-expressing plasmids pUC19RP12 and pST76-ASceP (accession numbers AF170481 and AF170483, respectively) were also developed in this laboratory and have been described (Pósfai et al. 1999).

3.2.2 Plasmids used in the scarless deletion method

Sequence data for plasmids pSG76-CS, pKSUC1, and pSTKST are available under accession Nos. AF402780, AF402779 and AF406953, respectively. The three plasmids were engineered from parental plasmids described previously (Pósfai et al. 1997, 1999). pSG76-CS serves as template plasmid to generate linear targeting fragments by PCR. It contains the Cm^R gene and two I-*Sce*I sites and was obtained by the PCR-mediated insertion of a second I-*Sce*I recognition site into pSG76-C, downstream of the *Not*I site. The two I-*Sce*I sites are in opposite orientation. pKSUC1 is a high-copy-number plasmid expressing I-*Sce*I constitutively. It was assembled by ligating the *Xba*I-*Not*I fragment of pSG76-K, blunted at the *Not*I-end and carrying the kanamycin resistance (Km^R) gene, to the *Xba*I-*Dra*I fragment of pUC19RP12, carrying the I-*Sce*I gene and the pUC origin of replication. pSTKST is a low-copy-number plasmid expressing I-*Sce*I under the control of the *tet* promoter-operator. It was derived by ligating the *Xba*I-*Pst*I fragment of pUC19RP12, carrying the I-*Sce*I gene, to the large *Xba*I-*Pst*I fragment of pFT-K, carrying the gene for Km^R. Replication of the plasmid is temperature sensitive and is abolished at 37°C to 43°C. pSG76-CSH was constructed by cleaving pSG76-CS by *Eco*RI and *Hha*I, blunting the ends by DNA polymerase I large fragment and recircularizing the plasmid by ligation. pBAD$\alpha\beta\gamma$ has been described (Zhang et al. 1998) and was a gift of A.F. Stewart (EMBL, Heidelberg).

3.3 PCR protocols

The PCR protocol, used for the generation of the linear targeting fragments, is described in detail in 3.6. For targeting fragment generation, high-fidelity Takara ExTaq polymerase was used. For other applications, Taq polymerase was obtained from Fermentas.

PCR screening of the colonies for identifying correct insertions was performed under the following conditions: 28 cycles: 94°C-40 sec/57°C-40 sec/72°C-80 sec. Briefly, colonies were resuspended in 20 µL TE buffer, and 2 µL of the suspension was added to a PCR mix (2,5 µL 10xPCR buffer, 2 µL of 2 mM $MgCl_2$ stock, 2 µL of 2.5

mM dNTP stock, 0,2 µL Taq polymerase, 16.3 µL H_2O) resulting in a total volume of 25µL. 5 µL of the PCR product was analyzed by agarose gel electrophoresis.

3.4 Electroporation

Electroporation-competent cells were prepared as described in the Invitrogen Electroporator II Manual (http://www.invitrogen.com). Briefly, a 500-ml *E.coli* culture was grown to an OD_{550} of 0.5-0.6, then cells were harvested by centrifugation and washed twice in ice-cold water and once in ice-cold 10% glycerol by repeated centrifugation and suspension. At the final step, the cell pellet was suspended in 0.4 ml 10% glycerol, aliquoted in 40 µl portions into microfuge tubes and stored at -80°C.

The cells were typically electroporated with nanogram quantities of plasmid DNA at 1.8 kV and a resistance of 150 Ω in a 0.1-cm electroporation cuvette using the Electroporator II device (Invitrogen). Cells were then diluted with 1 ml SOC medium, incubated in a shaker at an appropriate temperature (30°C- if cells harbor a ts plasmid, 37°C- if they have a non-ts plasmid) for 1 h, and plated on selective medium.

3.5 DSB-stimulated deletion procedure

Insertion of the targeting suicide plasmids into the chromosome

pST76-based suicide plasmids carrying a homologous segment to a selected region of the genome were delivered into the target cell by electroporation. Cells were spread on LB plates supplemented with an appropriate antibiotic (LB+Ab) and incubated at 30°C overnight. Several colonies were then picked and restreaked on a LB+Ab plate. This plate was incubated at 42°C for 7-9 h, then transferred to 37°C for an additional 12-24 h. Typically many large colonies are formed over the background of small colonies. These large colonies carry the suicide plasmid inserted into the chromosome; however, at this point usually the colonies contain cells harboring unintegrated plasmids as well. To obtain cells that are cured of the free plasmid, a few large colonies are picked individually, restreaked on LB+Ab plates and grown at 37°C overnight. Colonies should

be uniform in size this time and the site of insertion can be verified by PCR using an appropriate primer pair.

Insertion of pSG76-based suicide plasmids can be achieved in a similar way, except that the target cell should carry a temperature-sensitive replicon-based helper plasmid (pFT-A) supplying Π protein, and longer incubation time is needed at 42°C to stop replication of the helper plasmid and to dilute the Π protein in the growing cells.

Resolution of the cointegrate/generation of a genomic deletion

To express I-*Sce*I in the cell harboring a correct cointegrate, high copy-number pUC19RP19 is electroporated into the cell (alternatively, low copy-number pST76-ASceP can be used). The meganuclease introduces a DSB into the inserted plasmid sequence, stimulating resolution of cointegrate. The procedure can result in either wt sequence or deletion product, depending on the site of the crossover. Using a specific outside primer pair, cells carrying the desired deletion can be identified by PCR screening.

3.6 λ Red-mediated scarless deletion procedure

Construction of a linear targeting fragment by PCR

Steps of the procedure can be followed in Figure 3. To generate primer ab, 20 pmole of primer a was mixed with 20 pmole of primer b, and PCR was performed in a total volume of 50 μL. Cycle parameters were 15 × (94°C 40 sec/57°C (or lower depending on the extent of overlap between primers a and b) 40 sec/72°C 15 sec). Next, 1 μL of this PCR product was mixed with 20 pmole of primers a and c each and 50 ng of pSG76-CS template, and a second round of PCR was performed in a volume of 100 μL. Cycle parameters were 28 × (94°C 40 sec/57°C 40 sec/72°C 80 sec). The resulting PCR-generated linear DNA-fragment was purified by Promega Wizard PCR purification kit and suspended in 20 μL H_2O. Elimination of the template plasmid (e.g., by *Dpn*I digestion) is not needed.

Replacement of a genomic region by insertion of the targeting DNA-fragment

The target strain carrying pBADαβγ was grown at 37°C in the presence of Ap, and electrocompetent cells were prepared as described (Pósfai et al. 1999), except that 0.1%

L-arabinose was added to the culture 0.5 to 1 h before harvesting the cells; 4 µL targeting DNA-fragment (200 to 500 ng) was electroporated into 40 µL of electrocompetent cells. Cells were plated on Cm+Ap plates and incubated at 37°C. A total of 10 to several hundred colonies were usually obtained after 12 to 24 h of incubation. Typically, 12 colonies were checked for correct-site insertion of the fragment by colony PCR using primers d and e (Figure 3).

Deletion of the inserted sequences

A colony harboring the desired insertion was grown in LB+Cm, competent cells were prepared for chemical transformation, and pSTKST was introduced into them by standard procedures (Sambrook et al. 1989). Cells were spread on LB+Km plates and incubated at 30°C. A colony from this plate was inoculated into 10 mL of LB+Km supplemented with heat-treated inducer cTc (final concentration, 25 µg/mL) and grown for 24 h at 30°C. Dilutions of the culture were plated on LB+Km plates and incubated overnight at 30°C. Alternatively, instead of pSTKST, pKSUC1 can be used and the cTc induction step can be omitted. The transformation mixture is spread on LB+Km plates and incubated overnight at 37°C. Typically six colonies were screened by PCR using primers d and e. A fragment size matching the predicted length indicated the presence of the desired scarless deletion. Helper plasmids can be eliminated from the cells by growing the culture at 37°C to 43°C in LB. Replication of pSTKST is inhibited at this temperature, and pBADαβγ is rapidly lost from the cells under nonselective conditions.

3.7. Primers

Deletion	Primer sequence

MD1	Primers for MD1 deletion are described in Pósfai et al. 1999
MD2	a: 5’-gaccgacaagataccacgccagcaacatcagcagccataccattatgcagaaacaatg b: 5’-tcacgtactaagctctcatgtctgacgcaaaagcagccatcgacagcattgtttctgc c: 5’-ccagcgcactgcgcgataatcccaatgcacaatggaccagaaactcagaaggttcgtcc d: 5’[1398137]-gcagctgcaccatgcgttc e: 5’[1481227]-tggaaaatggcctgttgt
MD3	a: 5’-taaactgatatacattgccgaacaatgtggtctttagtttcataagtcgttccctcaggaaggaaatgcgagtgatttc b: 5’-tcacgtactaagctctcatggttgtgccggtgacggtgaaatcactcgcatttcctt c: 5’-cttcctgaatgcggtattcgccagcggtggcgctgggtggcagaaactcagaaggttcgtcc d: 5’[2556481]-cgtcacgcagtcatcttcat e: 5’[2564020]-cctgcgtgagatgaacgtaca
MD4	a: 5’-cttttaggcccattgataggcccaacgaaaagctctattgtttacgttgggcctaaacgcagggagactccccacctttga b: 5’-tcacgtactaagctctcatgaataagttactgtcgttaagtttcatcgctacatcctgttttcaaaggtggggagtctc c: 5’-tattgattaaatgcggcgtgtaaggcatccacactacgtgccagaaactcagaaggttcgtcc d: 5’[2754075]-cgaagtccgaagagaacta e: 5’[2789380]-gtcgatggcttcaccattgt
MD5	a: 5’-tttctcccgtaaatgccttgaatcagcctatttagaccgtttcttcgccatttaaggcgttatccccagtttgcagaaac b: 5’-tcacgtactaagctctcatggccgacagatgagttatgagcgcttttaatctcattacggagtttctgcaaactgggga c: 5’-gtgcgtgccgataagtcattaagcccgtttgaaatccgggtatacagaaactcagaaggttcgtcc d: 5’[2064207]-cttctcggctgactcagt e: 5’[2078795]-tgtgttaacgcgatgacca
MD6	a: 5’-tttgtcgggagaagataccagggtgacctgtttctttctgctcacgcgcggcgatatagaattcaaacattataatccct b: 5’-tcacgtactaagctctcatgtaaataattcctgcgtaggacttttgttttgcagtttttacgtcacaagggattataatgtttga c: 5’-atgaaattaaatatatttactaaatctatgattggtatggggcagaaactcagaaggttcgtcc d: 5’[3451385]-tactatccaaagcacccat e: 5’[3467730]-catttgttgacggcatatct
MD7	a: 5’-atcgttgacagccgcactccatgacgggtaaaaagtggataaaataatttttacccaccggatttttaccctagagccataca b: 5’-tcacgtactaagctctcatgatcaggatattttaagaaactgacaggcctcatcgagtgtgaggctgtatggctctagggtaa c: 5’-ttatttcacctcttgttgtgacattgttgaaaaatggataccagaaactcagaaggttcgtcc d: 5’[2464473]-agttcgtcatggttcgctt e: 5’[2474396]-caaagccagtatcagccat
MD8	a: 5’-aatccagtattacccgctcttaagcatcccgtgctatgttattgacacacaaaagcgttgaggaacagtgagaagcgcagtgcc b: 5’-cacgtactaagctctcatgaaagaagcactaaggatttggtgaaagcagcagtatcgggatggctgggcactgcgcttctcac c: 5’-gaatttatggatttcaagagtcatgcgtgttaactatttgatagaaactcagaaggttcgtcc d: 5’[1625425]-gtgttccatagcgtaggct e: 5’[1650955]-cattgctgcaaagtcttact
MD9	a: 5’-ttggtacagatgatggtatttccggttcgataatgcttgtaccaacagggagggaatacgagaaatcacaggacatt b: 5’-tcacgtactaagctctcatggtttttagcttcaggtaatattgcgtaccagcattagcaatgtcctgtgatttct c: 5’-aagtcacgccaataatcgattgcacattccctgcagtcacctgccctccggtagaaactcagaaggttcgtcc d: 5’[4493975]-cgaagtggcgaaatcggta e: 5’[4547758]-gctgcaatcctagatcca
MD10	a: 5’-ctcagcagcactgacatcaatttgtgtcacccgcagcgcatagattacgccattttgaa b: 5’-tcacgtactaagctctcatgtcactgataatgcgtccgttcgtaaattcaaaatggcgt c: 5’-aagcaatcaatggttttcatgacacggcgaatgcggtggcagaaactcagaaggttcgtcc d: 5’[3108503]-atctgacaataaccttcacg e: 5’[3134799]-acattcaggtatgtaccttc

MD11	a: 5’-agcttttatattgatttacaataagagtctcacctcctgattttagtaatctcttgatgctaagaactatcagataattc b: 5’-tcacgtactaagctctcatgtaaaaaacataattcctatcatcacagccacgctcattttgaattatctgatagttctta c: 5’-aaccgcaccacgaaaggtgcattgctaggcgcaggcgttggtttacttaccagaaactcagaaggttcgtcc d: 5’[1196246]-tccgtaaagatgttttcttcctg e: 5’[1222570]-tcgtacgcttatctttcgacaa
MD12	a: 5’- gcctgactgaatcagccagcttgcccaggtgtgtctgaggtcatggaaacggaaatcttcctcatttatgcccgt b: 5’- tcacgtactaagctctcatgtgatgtggtgaagggcaaagcggaaacggataagacgggcataaatgaggaaga c: 5’- tgcggcccacgacttagaagttcctagaacgacattttaagtcaaagaaactcagaaggttcgtcc d: 5’[563961]- ttggatagagcaacgacct e: 5’[585404]- atctcgccaaacagcgaca
MD13	a: 5'-tagggtgcgttgaagatatgcgagcacctgtaaagtggcggggatcactctacctcaatgtgtatcacaa b: 5'-tcacgtactaagctctcatgctccagactcccccacaaagaatatggatattgtgatacacattgaggta c: 5'-ggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtagaaactcagaaggttcgtcc d: 5'[15289]-gacccgctaacctcccca e: 5'[20662]-atcgtagaaccatcggctg
MD14	a: 5'-ttactgcttaaatgcacccgccagagagcgaatatcattgccggttggcgagtaatccccgcatatcc b: 5'-tcacgtactaagctctcatgagcaggtgccttaccatcctgacctgacaaccggatatgcggggattact c: 5'-cccataagcgctaacttaagggttgaaccatgtgaagaatgcgacgcctagaaactcagaaggttcgtcc d: 5'[602558]-aggcattcacgctacatcc e: 5'[608673]-atctggtacgattcgcc
MD15	a: 5'-tttgcaagctgggcaaactaagtatctgaccccgcataaggaatagaacacaggcggcactaacccggac b: 5'-tcacgtactaagctctcatggctttttgccgctttgtaatacagtctacgtccgggttagtgccgcctg c: 5'-gtgaaagcatcattggatgaaaaatcggcaacaggctggccccctgtttagaaactcagaaggttcgtcc d: 5'[2507553]-ggtcacacaattactttatcgt e: 5'[2516055]-ccaaccctcgctgcaaagc
MD16	a: 5'-agaaggaatcagaatttccaggtcagacgggctgcaagttgcagaccgttacgagagtggacggtccc b: 5'-tcacgtactaagctctcatgctcaccctaaccctctccccaagggggcgaggggaccgtccactctcgta c: 5'-ttaccccgcccgataaaacggtttatccccaaggatcgtcgcccgatgcagaaactcagaaggttcgtcc d: 5'[379203]-atccggtgaatgtggtcg e: 5'[387953]-gagttaacctgttgaaagta
MD17	a: 5'-ttaatagaaacagcaataattttatattcactgaaaatatttttaatcttgaatattatcgggcgttaa b: 5'-tcacgtactaagctctcatggataacgcattgaaaaccacaatgtaatattttaacgcccgataatattc c: 5'-ttattgggcgttttcagtttcttcatcgcgacgacgaataatcaaccccagaaactcagaaggttcgtcc d: 5'[389023]-tgtcgcgatgttgacaaga e: 5'[399128]-cgatctgatcacacgtttg
MD18	a: 5'-gtagatttgttacactccgctatagctttcttcagatctttatagcccacttatttaacaacttga b: 5'-tcacgtactaagctctcatgttgtaacaacccaggtaaaacaccttattttcaagttgttaaataagtgg c: 5'-gttatctacgcttttaacaaacgattgatggaatattttatgaaagggagaaactcagaaggttcgtcc d: 5'[2992875]-atgagcatctatcttatgcga e: 5'[2996992]-gtaatacttccccattttgt
MD19	a: 5'-ctaatgattagcggcagcaacgcatagcttcacataattctggtttatatctaatacaaacactaaa b: 5'-tcacgtactaagctctcatgagcactgatataacgggcctgatggcccgttttagtgtttgtattagata c: 5'-ctagttataggtgattttcaggattacgtctttactaaacggtccctcagaaactcagaaggttcgtcc d: 5'[3182723]-ctgagaataagcggattcac e: 5'[3189811]-tatgcctggtattgtcaaca
MD20	a: 5'-atcgtctgcaactttattgtgcagtgttgtgcctgttagggaaggtgccttaggtaacatttagtttgg b: 5'-tcacgtactaagctctcatgacaataaaacagcaatatctttacatttagccaaactaaatgttacctaa c: 5'-ggaaggtgcgaacaagtccctgatatgagatcatgtttgtcatctggagagaaactcagaaggttcgtcc d: 5'[1386813]-aactcctttatcttctgcgt e: 5'[1396740]-ttgcttcggaatggtgcg

MD21	a: 5'-aacgtcacacctagtttatgctaactgtcaataacacagcaaacgctattacagggtctccttacaacaa b: 5'-tcacgtactaagctctcatgtaatcttttaattacaaactgcgttgacagttgttgtaaggagaccctgt c: 5'-atggctatgtatcagaacatgctcgttgttatcgatcctaaccaggacgagaaactcagaaggttcgtcc d: 5'[686975]-tggaaatgaacggtttgtga e: 5'[688355]-atctggcagaaggttcatca
MD22	a: 5'-aagtataaatagcttatccatgcttatatgcttacggctttatatgtctccacatgcgatatttc b: 5'-tcacgtactaagctctcatgtaaaagtggtaataataaaacattatttaagaaatatcgcatgtggagac c: 5'-actatattaaatattgtcaacctaaagaaactcctaaaaaccatattgagaaactcagaaggttcgtcc d: 5'[2098958]-acgacgaatagtgtcctgga e: 5'[2135819]-gagggtaataacctcacatt
MD23	a: 5'-cagaccgtgacacatcacagcctgtttatttttctgttatcagaacgtcctgagctttcatcctatgaaa b: 5'-tcacgtactaagctctcatgtctgtacccaatgcttttaacagcaattaatttcataggatgaaagctca c: 5'-atgcatcaatctggttctgtttctctttgtcgttccgcaatatctgttcagaaactcagaaggttcgtcc d: 5'[2284221]-agccacgagtcggcacgt e: 5'[2288281]-actgactccgggtatgga
MD24	a: 5'-atgctcagacacattacattcactgtatttataacaacatcaatgaatactaacaatgccggacaggagt b: 5'-tcacgtactaagctctcatgaccgatcgtataaaacgaatgcctgatagtactcctgtccggcattgtta c: 5'-tcacaagtagactacgcttatcacagctttcccatcaagttcgatttttagaaactcagaaggttcgtcc d: 5'[3359650]-tggtctgtcttaatcaatctc e: 5'[3365359]-caagatatgcgaagagctttc
MD25	a: 5'-atccaatttaaacatcttagcataaaacaaatgatgaataaggaattttatattgattattagcacttta b: 5'-tcacgtactaagctctcatgtacataaaatctatttttattcaatgattataaagtgctaataatcaata c: 5'-ttatggtattatattctggtaatgatttgataagcgaattcttttaccagaaactcagaaggttcgtcc d: 5'[3648836]-attgtgtgtagaagaaaatc e: 5'[3651443]-gcgtgacatgaatatccat
MD26	a: 5'-aagacatcgagccttttttcactgagttattaaacatactcgcgagcgcgtcgctacgctagcaagc b: 5'-tcacgtactaagctctcatgaacatcccatttcgattattcctgtttcatttttgcttgctagcgtagcg c: 5'-taacagattgaaatacacccaaaacaaaagtatgacttatacatttatgagaaactcagaaggttcgtcc d: 5'[1960419]-actcgccagttgcagctat e: 5'[1977512]-tagtttgttgtgcggtaag
MD27	a: 5'-atgcgtagaccggaaaaggcgttcacgccgcatccggccactttcagtttgacaatacgtacgcttgaga b: 5'-tcacgtactaagctctcatggatgttatactcaaagtagacaatcctgatctcaagcgtacgtattgtc c: 5'-ggtaatgactccaacttattgatagtgttttatgttcagataatgcccgagaaactcagaaggttcgtcc d: 5'[1041154]-attcagttacctgaatgctc e: 5'[1049868]-ctacccgcagtatatagtgc
MD28	a: 5'-gtgtgttatcggtgcagagcccgggcgaaccgggctttgttttgggtgttcccggtgtgtgcaatatac b: 5'-tcacgtactaagctctcatg tcgtacgaaagtacgggctttttttttcgtatattgcacacaccggg c: 5'-ttcgacaactggcgacagccagttggacagaccgtgaacgcagtgagcgagaaactcagaaggttcgtcc d: 5'[1085198]-aatggttgggatcgtcc e: 5'[1096861]-gtggctgaaggagcacg
MD29	a: 5'-caggcatcagaaaatcgctcatactttaatcggtaaacagcacctttagaatcctgtcgtttgttttcga b: 5'-tcacgtactaagctctcatgaaaaacaaataatttcattatattttgaaatcgaaaacaaacgacaggat c: 5'-atgaaaatgtacgtggtatcgacaaagcagatgctgaacaacgcacagcagaaactcagaaggttcgtcc d: 5'[2163073]-aacattatcccacgcatggt e: 5'[2175378]-actttgctacggcttcccta
MD30	a: 5'-gaaagtcaatgatgttgaaagggacatttaccccaaagaggacaaaggaacaaaatagccctcttcccac b: 5'-tcacgtactaagctctcatggcggggtactaaggttagcggccctcttcgtgggaagagggctattttg c: 5'-tcagggcaagccgaaggtagccttttttatttcgtaatcctgtagatatagaaactcagaaggttcgtcc d: 5'[3578670]-ttcgcacttgtccaacact e: 5'[3582808]-gatgggcagtacacaaagca

MD31	a: 5'-ctatatttctgcggcattcccagcctcctgtgttgatttccaacgagtataactttgatttatagtcagg b: 5'-tcacgtactaagctctcatgtcaccgaaaggcagacagagaaaagccccacctgactataaatcaaagtt c: 5'-acgaggcatataggtctgattccgatattcaattggagtcagacctagaaactcagaaggttcgtcc d: 5'[3718018]-gtgcaatgagaatgcgcaa e: 5'[3719782]-ggtgcttgagactgtttgt

Note: MG 1655 genomic coordinates of the 5' nucleotides of primers **d** and **e** are shown in **[]**.

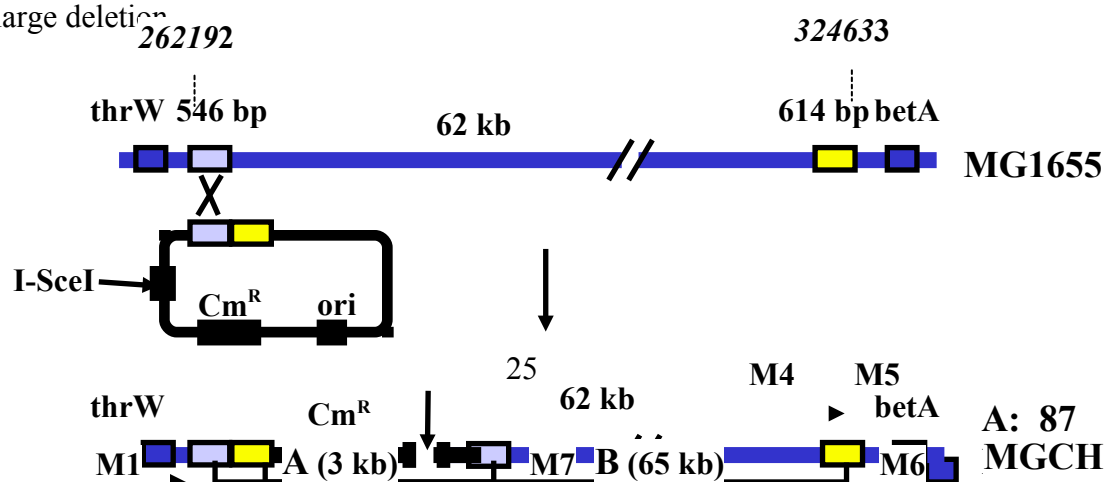
4. Results

4.1 Deletion of a 62-kb genomic segment by the DSB-stimulated method

Earlier work in the laboratory indicated that the suicide plasmid-based, DSB-stimulated method could be used to generate targeted deletions in the genome. A joint fragment of two genomic segments, which flank the planned deletion, is cloned in the targeting plasmid, and recombined into the genome. Insertion of the plasmid creates a duplication of one of the cloned segments in the chromosome. DSB-stimulated recombination of the duplicated segment pair results in the desired deletion. However, due to the topology of cointegrate, large deletions are difficult to obtain this way. Since the frequency of intramolecular recombination between direct repeats depends on the physical distance on the chromosome (Lloyd et al. 1996), resolution of the cointegrate would preferentially lead to reversion to wt. In a preliminary experiment, using *sacB*/sucrose counterselection, short-distance deletion leading to wt was heavily favored over a long-distance deletion (46 kb), which could not be detected by PCR in a pool of several hundred colonies (data not shown).

To assess the applicability of the DSB-stimulated method in generating large deletions, two variations of the procedure were tested and compared in an experiment aimed at deleting a large K-island of the MG1655 genome. While the first version involves integration of a single suicide plasmid in the genome, the second version is based on two separate insertion events.

To test the efficiency of the two versions, pSG76-CMB45 was inserted into the genome of MG1655 (Figure 4), resulting in cells designated MGCH. The plasmid carried PCR-amplified copies of two, relatively short segments (546 and 614 bp) flanking a 62-kb chromosomal region that coded for putative prophages and was assumed dispensable. The I-*Sce*I-expressing plasmid pUC19RP12 was electroporated into MGCH, and surviving transformants were tested by PCR, using primers M1/M6 and M1/M7. Of 100 colonies checked by PCR using primers M1 and M6, 87 reverted to wt and 13 contained the large deletion.



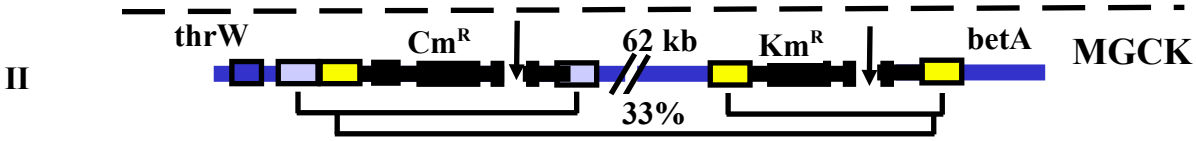


Figure 4. Construction of a 62-kb deletion of the MG1655 chromosome. (I) Deletion of the 62-kb segment using a single insertion. The genomic coordinates of the 5' nucleotides of the homologous segments are shown in italics. Arrowheads marked M1, M7, M4, M5 and M6 are PCR primers. Type A recombination event, found in 87 of 100 colonies, deletes the suicide plasmid; type B deletion, found in 13 of 100 colonies, removes the large, 62-kb chromosomal region. (II) Deletion of the 62-kb segment with the help of two insertions. DSB-stimulated resolution of the cointegrate results in the loss of the chromosomal segment in 33% of the cells (average of three experiments). I-*Sce*I was expressed either from pUC19RP12 or from pST76-ASceP.

As a more effective but more laborious alternative, a second insertion, using pST76-KM45, was introduced into the genome of MGCH, resulting in strain MGCK (Figure 4). Plasmid pUC19RP12 was electroporated into MGCK. I-*Sce*I expression causes two DSBs, one in each of the insertions. Based on three independent experiments, a small

number (0.5-1%) of the potential transformants survived, but an average of 33% of the cells carried the desired deletion. (The number of potential transformants was determined by electroporating MG1655 and MGCK cells with both pUC19RP12 and pUC19 control plasmid in parallel experiments.) In a similar experiment, using low copy-number I-*Sce*I-expressing plasmid pST76-ASceP, nearly 100% of the transformants survived, and the fraction of the deletion mutants varied between 6 and 18% in three independent experiments.

4.2 Development of a λ Red-mediated scarless deletion procedure

To construct a series of targeted deletions, a more straightforward and more rapid method than the suicide plasmid-based, DSB-stimulated procedure was needed. By combining the appropriate features of the DSB-stimulated and the λ Red-mediated procedures, a new scarless method was designed.

Outline of the deletion method

Steps of the method are depicted in Figure 3 and detailed in *Materials and Methods.* To delete the chromosomal region between arbitrarily chosen segments (indicated as box A and B), first a linear DNA molecule is generated by PCR on the template plasmid pSG76-CS. The resulting fragment, carrying a selectable marker (chloramphenicol resistance, Cm^R) flanked by two I-*Sce*I meganuclease sites, is electroporated into the target cell where it can replace a segment of the chromosome. This replacement requires a double crossover, and involves short (40-60 bp) terminal segments ("homology boxes" **A** and **C**) of the DNA fragment. The helper plasmid pBADαβγ (Muyrers et al. 1999) provides the arabinose-inducible recombinase functions (Red αβγ) necessary for the integration event. Cells which integrated the linear fragment into their chromosome are then selected by their Cm resistance (hereafter referred to as "intermediate" clones). Correct-site insertions are confirmed by PCR screening using primers d and e. The procedure is

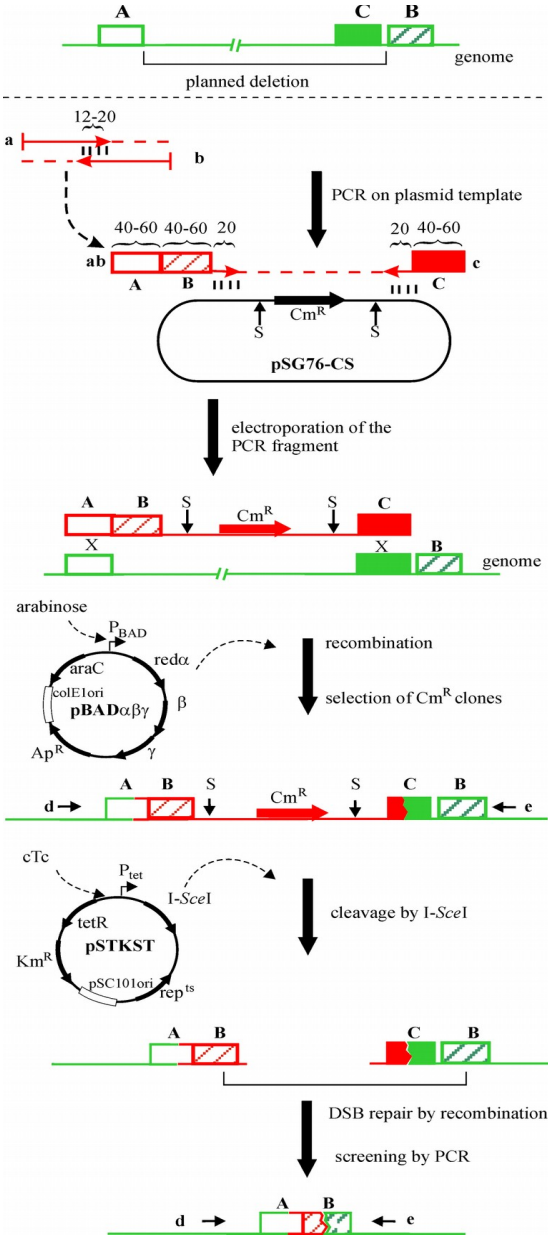


Figure 3. Overview of the deletion procedure. A, B, and C represent arbitrarily chosen DNA segments (homology boxes). Numbers refer to the length of oligonucleotides in basepairs. Polymerase chain reaction primers are labeled by lower case letters in bold (**a**, **b**, **c**, **d**, **e**, **ab**). *S* indicates an I-*Sce*I cleavage site. The 3' ends (~17-20 bp) of primer **c** and complex primer **ab** are complementary to the template plasmid pSG76-CS upstream and downstream of the Cm^R gene.

similar to published protocols (Zhang et al. 1998; Datsenko and Wanner 2000; Yu et al. 2000). However, in contrast to the original λ Red recombination method, the integrated fragment carries a third homology region (box B) fused to box A as part of the PCR primer. Since synthesis of very long primers is difficult, the fusion is produced in a PCR-like filling-in reaction of two partially complementary oligonucleotides (primers a and b), resulting in the composite primer ab, which in turn is used to initiate the generation of the linear DNA fragment. Note that the bulk of the deletion is generated by the integration of the fragment. Moreover, the sequence corresponding to box B is duplicated in the chromosome at this stage.

Next, I-*Sce*I meganuclease expression is induced by derepressing the *tet* promoter on the helper plasmid pSTKST. (Alternatively, pKSUC1, a plasmid constitutively expressing I-*Sce*I, can be transformed into the cell.) As a result, the chromosome is cleaved at the 18-bp I-*Sce*I sites (Monteilhet et al. 1990) present on the integrated fragment. The broken chromosomal ends are then subjected to RecA-mediated DSB repair by intramolecular recombination. Since the broken ends carry short homologous regions (box B) close to their termini, recombinational repair is likely to proceed *via* these homologous segments. Surviving colonies are screened by PCR using primers d and e. Correct-size PCR fragments confirm the generation of the desired, scarless deletion.

Features of the deletion method

Several large deletions have been constructed using the new method (Table I). Total length of the targeting linear DNA fragment was 1.7 kb. Sizes of the homology boxes A, B and C were in the range of 40-80, 40-60 and 40-45 bp, respectively. Typically 200 to 500 ng DNA fragment was electroporated into MG1655 and 10 to 200 recombinant (Cm^R) colonies were obtained, 5 to 90 % of which contained an insertion in the desired locus. In our hands, this relatively low number of correct insertions was the limiting factor in the procedure. There were indications in the literature that internal regions of the fragment homologous to chromosomal sequences cause a drop in the integration efficiency in λ Red-mediated insertion procedures. The template plasmid pSG76-CS carries a 78-bp piece of IS1 that is copied into the targeting DNA fragment. Since IS1 is present in 7 copies in the MG1655 genome, it seemed likely to interfere with the

integration process. However, when the 78-bp internal homology was deleted from the plasmid, and the resulting pSG76-CSH was used as a template to generate the targeting fragment, no increase was observed in the number of integrants.

Sizes of the deletions were in the range of 7 to 82 kb. No significant correlation was found between the efficiency of integration and the size of the deletion it caused. In the case of deletion MD2, two different targeting fragments were constructed. The DNA fragment replacing a 2-kb segment of the chromosome produced five times more integrants (53 colonies) than the one replacing an 82-kb segment (10 colonies). However, it must be noted that the results are not fully comparable, since the box C segments were from different genomic regions in the two constructs.

The second step of the procedure, DSB-stimulated recombinational repair proceeded efficiently. Following induction of I-*Sce*I expression, 10-100 % of the surviving cells carried the desired deletion. However, initial experiments showed that positioning the broken chromosomal ends relatively close to box B segments is essential, otherwise recombinational gap repair can proceed *via* other, randomly occurring, short homologies. Therefore, it is advisable to select a box C sequence close to box B on the chromosome. For similar reasons, in an experiment involving deletion MD2, placing two I-*Sce*I sites on the targeting fragment (as shown in Fig. 3) resulted in an increased level (15 %) of recombinational repair proceeding via box B, as compared to having a single I-*Sce*I site on the fragment (5 %).

Deletions were created in the recombination-proficient strain MG1655. A number of recombinational activities are thought to be involved in DSB repair (reviewed in Kowalczykowski 2000), with RecA playing a central role. Since recombination of short (40-60 bp) repeats, atypical substrates for the recombinase, are involved in the deletion process, the role of RecA was tested. It was found that deletion of the *recA* gene practically abolished recombinational repair when pUC19RP12, a plasmid constitutively expressing I-*Sce*I, was transformed into MG1655 cells carrying the insertion corresponding to the intermediate construct for deletion MD12 (data not shown).

4.3 Deletion target selection

Chromosomal regions marked for deletion were chosen with the expert help of our collaborator Guy Punkett III (University of Wisconsin). Selection was primarily based on comparing the genomes of MG1655 and EDL933. We hypothesized that the regions present only in one of the two strains are unlikely to encode essential functions under common culturing conditions. In order to maximize the amount of DNA removed, 12 large genomic regions, including the 11 largest and 6 smaller K-islands, were selected for deletion (Table 1, Figure 5). Candidate regions were checked for the presence of potentially essential genes by searching databases. In four cases (MD1, MD2, MD4, MD9), relatively large regions (6-15 kb) adjacent to the K-islands and encoding several genes of unknown functions were included in the target. Generally, deletion endpoints were placed in non-coding regions, next to the nearest ORFs to be deleted. Due to primer design considerations, in a few cases the deletion endpoint was shifted into coding sequences leaving the 3'-end region of the deleted ORF in the genome (Table 2). No new fusion protein was created by the deletions. Eight of the 12 deleted regions carried cryptic prophages or phage remnants. All together, the segments marked for deletion possessed a higher than average ratio (50% versus 38%) of ORFs classified as unknown (Blattner et al. 1997).

4.4 Deletions

A total of 12 regions (MD1-MD12) were sequentially deleted from the MG1655 genome. The strain carrying all 12 deletions was designated MDS12. Coordinates of the deletions are shown in Table I. A total of 376,180 nucleotides were deleted, resulting in an 8.1 % reduction in the size of the genome. All together, 409 protein-coding ORFs and 2 stable RNA genes were removed, 9.3 % of the total gene count.

Deletion	Endpoints [a]	Size (bp)	Description [b]
MD1	263080, 324632	61553	b0246-b0310; includes K-islands #16, 17, 18, CP4-6, eaeH
MD2	1398351, 1480278	81928	b1336-b1411; includes K-island #83, Rac
MD3	2556711, 2563500	6790	b2441-b2450; includes K-island #128, CP-Eut
MD4	2754180, 2789270	35091	b2622-b2660; includes K-island #137, CP4-57, ileY
MD5	2064327, 2078613	14287	b1994-b2008; includes K-islands # 94, 95, 96, CP4-44
MD6	3451565, 3467490	15926	b3323-b3338; includes K-islands #164, 165
MD7	2464565, 2474198	9634	b2349-b2363; includes K-island #121
MD8	1625542, 1650785	25244	b1539-b1579; includes K-island #77, Qin
MD9	4494243, 4547279	53037	b4271-b4320; includes K-island #225, fec operon, fim operon
MD10	3108697, 3134392	25696	b2968-b2987; includes K-island #153, glc operon
MD11	1196360, 1222299	25940	b1137-b1172; includes K-island #71, e14
MD12	564278, 585331	21054	b0538-b0565; includes K-island #37, DLP12

Table 1. Genomic deletions in strain MDS12.

[a] Endpoints indicate the first and last nucleotides in the genome of MG1655 that were removed in each deletion.

[b] Descriptions of the deletions include the span of genes deleted (indicated by b-number; Blattner et al. 1997), the MG1655-specific regions (K-islands; Perna et al. 2001), and specific elements such as genes, operons, and cryptic prophages.

Deletion	Sequence left of endpoint	Sequence right of endpoint
MD1	tgggggctgagttcaccggtggtcagcatcgattccatc tggaggataaagtgcgggaat	atagtctgtatcaggaatgttcgggttaaatatcagcaaaaagcc cgcatcatgaatact
MD2	tacccggacctatcggcagaccgacaagataccacgc cagcaacatcagcagccatacCA	Ttatgcagaaacaatgctgtcgatggctgcttttgcgtcagactg tgctttcgctgccat
MD3	catccttaaactgatatacattgccgaacaatgtggtcttt agtttcataagtcgttccc	tcaggaaggaaatgcgagtgatttcaccgtcaccggcacaacc gatccgccaaaaagagg
MD4	tgataggcccaacgaaaagctctattgtttacgttgggcc taaacgcagggagactcccc	acctttgaaaacaggatgtagcgatgaaacttaacgacagtaact tattccgccagcaggc
MD5	atgccttgaatcagcctatttagaccgtttcttcgccattta aggcgttatccccagttt	gcagaaactccgtaatgagattaaaagcgctcataactcatctgt cggccagaaggcgcg
MD6	gaagataccagggtgacctgtttctttctgctcacgcgcg gcgatatagaattcaaacat	tataatcccttgtgacgtaaaaactgcaaaacaaaagtcctacgc aggaattatttacgt
MD7	gccgcactccatgacgggtaaaaagtggataaaataatt tttacccaccggatttttaccc	tagagccatacagcctcacactcgatgaggcctgtcagtttctta aaatatcctgatcta
MD8	ccgctcttaagcatcccgtgctatgttattgacacacaaa agcgttgaggaacagtgaga	agcgcagtgcccagccatcccgatactgctgctttcaccaaatc cttagtgcttctttcg
MD9	ttggtacagatgatggtatttccggttcgataatgcttgtac caacagggagggaatacg	agaaatcacaggacattgctaatgctggtacgcaatattacctga agctaaaaacctgca
MD10	attcaccatcccaggcttctcagcagcactgacatcaatt tgtgtcacccgcagcgcata	gaTTacgccattttgaatttacgaacggacgcattatcagtgact ttaacggcatgggca
MD11	ttgatttacaataagagtctcacctcctgattttagtaatctc ttgatgctaagaactat	cagataattcaaaATGAgcgtggctgtgatgataggaattatg ttttttacgtgaatgag
MD12	gcctgactgaatcagccagcttgcccaggtgtgtctgag gtcatggaaacggaaatcttc	ctcatttatgcccgtcttatccgtttccgctttgcccttcaccacatc acttttgtcgc

Table 2. Predicted/empirical nucleotide sequences of the chromosomal joints formed by the deletions.

Notes:

1. In three cases (MD2, MD10, MD11) empirical data differ from the predicted sequence, due to small deletions affecting nucleotides shown in capitals. These small deletions are presumably caused by misannealing of primer a or b.
2. For the actual genome sequence delete nucleotides in capital letters.
3. There is a 114-bp pSG76-C/pST76-K vector segment between the left and right side sequence (*frt* sequence in bold) of MD1:

left – tccggtctccctatagtgagtcgtattaatttcgataagccagatcaagcttgtcgaccaattc

gaagttcctatactttctagagaataggaacttccgggatcaattcatcg - right

Deletion MD1 was created using a suicide plasmid-based, Flp recombinase-assisted method, described earlier (Pósfai et al. 1997). As a consequence, a 114-bp vector fragment, including a FRT site, remained in the chromosome. Since the *frt* sequence could serve as a convenient entry site for various insertions, no attempt was made to reengineer this deletion to remove the extra nucleotides.

Deletions MD2 through MD12 were constructed by the scarless method. However, as the work progressed, several variations of the procedure were tested. MD2 through MD6 were made in the same strain sequentially, and the incompatible helper plasmids pBADαβγ and pKSUC1, possessing different antibiotic resistance markers, were introduced into the strain by repeated alternating transformations to replace each other in a cyclic fashion. To construct deletions MD7 through MD12, I-*Sce*I was expressed from pSTKST. Some insertion intermediates were first constructed in wild-type MG1655 carrying pBADαβγ . The inserted regions were then transduced by P1 phage (Miller 1992) to the strain carrying all previous deletions and the second step of the procedure, DSB-stimulated recombinational repair, was carried out in this final host.

All 12 new chromosomal joints, formed by the deletions, were sequenced across (Table 2). Test primer pairs (d, e) were used to generate a PCR fragment spanning the deletion, and the fragment was sequenced from the ends. A minimum of 60-bp sequences were determined both upstream and downstream of the joint. In most cases results matched the predicted sequence. In three cases, however, small unintended alterations (a 4-bp, a 3-bp and a 2-bp deletion at MD11, MD2, and MD10, respectively) occurred at the 3'-ends of primer a or b, and were likely caused by misannealing of the primer ends in the filling-in reaction.

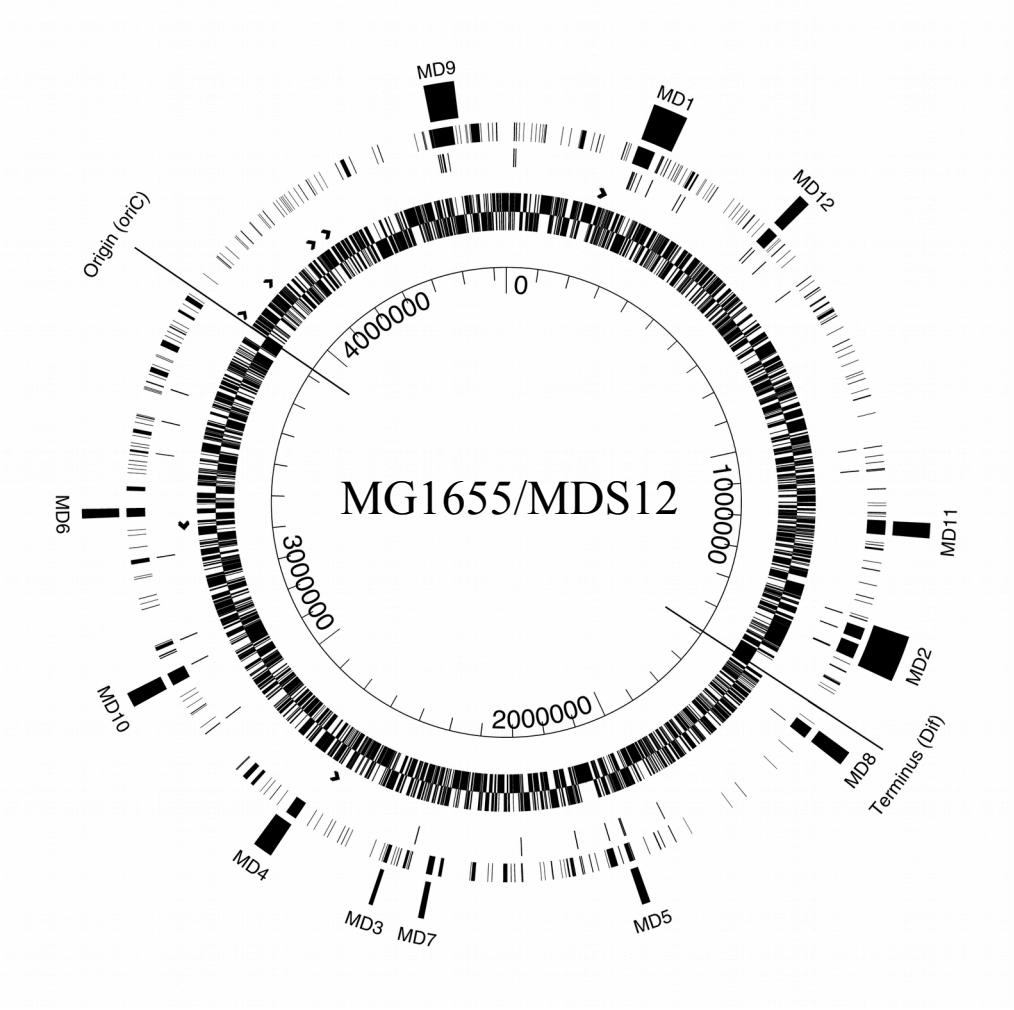


Figure 5. Representation of deletions MD1 through MD12 on the circular map of the MG1655 genome. Concentric rings denote the position of (from outside to inside) (1) deletions, (2) K-islands, (3) insertion sequence elements, (4) rRNA operons, and (5, 6) genes, by orientation. The replication origin and terminus are indicated.

4.5 Characterization of MDS12

Growth rates of MDS12 and the parental strain MG1655 were compared in both LB and M9/glucose minimal medium (Figure 6). Cultures were grown in duplicate and experiments were repeated twice. The doubling times of the two strains proved to be identical at exponential growth phase (33 min in LB, 41 min in minimal medium). A small but detectable difference in culture densities was observed in stationary phase. In both LB and minimal medium, MDS12 reached an OD ~10 % higher than MG1655.

Transformability of MDS12 and MG1655 was measured by electroporating a small, supercoiled plasmid, pTZ18U (Bio-Rad) into the cells. Electroporation efficiencies for the two strains were identical ($4x10^8$/μg plasmid DNA). Similarly, chemical transformation (Sambrook et al. 1989) efficiencies of the two strains were identical ($5x10^5$/μg plasmid DNA).

Additional characterization, including testing for phenotypic changes under a number of growth conditions, was done by other members of our laboratory and by our collaborators in the U.S.

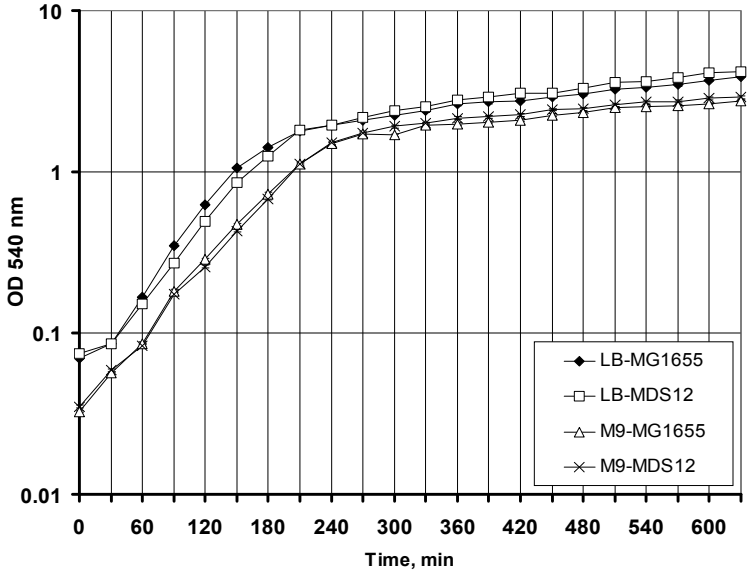


Figure 6. Growth curves of MG1655 and MDS12. Cultures were grown in duplicate in either LB or M9/glucose minimal medium. One hundred mL of liquid medium was inoculated with 1 mL overnight starter culture and grown at 37°C with moderate aeration.

4.6 Additional deletions

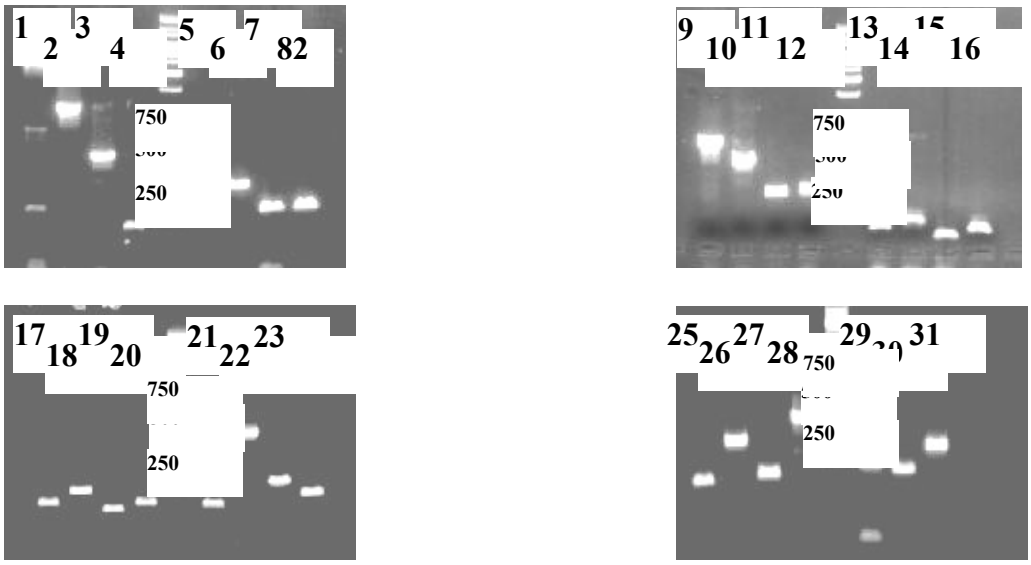
In the next phase of the *E. coli* genome remodeling work, an additional set of 19 deletions was constructed (Table 3). New deletion targets were selected to remove all ISs from the genome. In some cases, target segments were expanded to include dispensable regions adjacent to the ISs, resulting in the deletion of additional K-islands, ORFs with unknown functions, and the O-antigen and colanic acid gene clusters.

All deletions were made in two steps. Insertion intermediates were first constructed in wild-type MG1655 carrying pBADαβγ . The inserted regions were then transduced by P1 phage (Miller et al. 1992) to the strain carrying all previous deletions and the second step of the procedure, DSB-stimulated recombinational repair, was carried out in this final host.

The strain carrying all 31 deletions was designated MDS31. The presence of all deletions in MDS31 was confirmed by PCR using corresponding primers d and e (Figure 7). Characterization of the strain is currently being carried out by other members of the laboratory. Preliminary results show that the growth characteristics are identical to that of the wt strain. Mutation rate in MDS31 appears to be lower than in MG1655, but results vary depending on the test system used.

Strain designation	Deletion endpoints	Deletion size (bp)
MDS13	15388, 20562	5175
MDS14	602688, 608572	5885
MDS15	2507651, 2515959	8309
MDS16	379334, 387870	8537
MDS17	389122, 399029	9908
MDS18	2993014, 2996890	3877
MDS19	3182797, 3189712	6916
MDS20	1386912, 1396645	9734
MDS21	687083, 688267	1185
MDS22	2099418, 2135738	36321
MDS23	2284421, 2288200	3780
MDS24	3359797, 3365277	5481
MDS25	3648921, 3651342	2422
MDS26	1960590, 1977353	16764
MDS27	1041254, 1049768	8515
MDS28	1085330, 1096545	11216
MDS29	2163173, 2175230	12058
MDS30	3578769, 3582673	3905
MDS31	3718263, 3719704	1442

Table 3. Coordinates and sizes of additional 19 genomic deletions.



Deletion	Predicted size (bp)	Deletion	Predicted size (bp)
MD1	2595	MD17	197
MD2	1162	MD18	240
MD3	749	MD19	172
MD4	214	MD20	193
MD5	301	MD21	195
MD6	419	MD22	540
MD7	289	MD23	280
MD8	286	MD24	228
MD9	746	MD25	185
MD10	600	MD26	329
MD11	384	MD27	199
MD12	389	MD28	447
MD13	198	MD29	247
MD14	230	MD30	233
MD15	193	MD31	322
MD16	213		

Figure 7. The presence of all deletions in MDS31 was confirmed by PCR using corresponding primers d and e. Lanes 1 to 31 of the gel picture correspond to deletions MD1 to MD31. The sizes of the PCR fragments match the predicted sizes shown in the table. Marker ladder is GeneRuler™ 1kb DNA Ladder from Fermentas. Some marker fragment sizes are indicated in the figure.

5. Discussion

5.1 Genome engineering methods

Earlier methods of genomic engineering are based on the use of suicide plasmids. Equipped with a proper targeting fragment, these suicide plasmids can integrate into the genome in a RecA-mediated homologous recombination process, forming a cointegrate structure. Spontaneous resolution of cointegrate can result in a targeted modification of the genome. The major difficulty of this approach is that counterselection procedures, needed for elimination of cells that have not undergone cointegrate resolution, are strain-, medium- or temperature-dependent, limiting the use of the method.

Recently developed methods for genomic engineering utilize λ bacteriophage recombinases and are based on homologous recombination of a PCR-generated linear DNA fragment carrying a selectable marker gene into the genome (Zhang et al. 1998, Datsenko and Wanner 2000, Murphy et al. 2000, Yu et al. 2000). In applications aimed at deletion construction, the inserted fragment replaces the chromosomal region marked for deletion. The method is fast, however, removal of the marker gene requires a second step. One option is to use an FRT- or loxP-flanked selection marker gene that can subsequently be removed by site-specific recombinases. The disadvantage is that there still remain some foreign sequences (including FRT or loxP sites) in the genome and they may interfere with gene expression or with subsequent rounds of deletions. Another solution is to replace the inserted sequences by a PCR-generated, markerless fragment in a second round of recombination. This procedure, however, requires that the original insertion carries a counter-selectable marker that is usually strain- and culturing conditions-specific.

The methods presented here apply I-*Sce*I-inflicted DSB of the chromosome as a counterselection tool. Meganuclease cleavage has a dual role: it (i) stimulates the intended recombination by producing free DNA ends and (ii) provides selection pressure for the loss of the inserted sequences responsible for deleterious chromosomal breaks. Since cleavage of the chromosome is not growth condition- or strain-specific, it is practically universally applicable.

First, we explored the applicability of the suicide plasmid-based, DSB-stimulated method in the genome reduction work. Both variations of the method (single insertion and double insertion) were suitable for efficient generation of deletions. However, this method requires time-consuming cloning steps, which is a serious disadvantage when a high number of consecutive deletions are planned.

Next, merging the DSB-stimulated recombination with the λ Red-mediated, linear DNA-based genome engineering method, a new procedure was developed. It combines the speed of λ Red-type recombination procedures with precise and efficient removal of all foreign sequences. The bulk of the chromosomal sequences to be deleted are replaced by inserting a PCR-generated fragment into the genome by λ Red-type recombination. The novelty of this step is that the inserted fragment carries the actual joint of the final deletion in its synthetic primer part and, consequently, 40-60-bp direct repeats, separated by about 1.7 kb, are created on the chromosome. Intramolecular recombination between the duplicated sequences then results in the loss of the selectable marker gene and all foreign sequences producing a clean, scarless deletion. However, spontaneous recombination of the short, duplicated sequences is a rare event. To stimulate recombination, in the second step of the procedure I-*Sce*I meganuclease is expressed, causing DSBs at the two I-*Sce*I sites uniquely found on the inserted fragment. In surviving cells, DSBs are repaired primarily by the desired intramolecular homologous recombination of the duplicated sequences positioned close to the broken ends. We note that in some cases intramolecular recombinational repair of the DSBs can proceed via other, randomly occurring short homologies flanking the site of damage. In fact, in some cases the majority of the surviving cells displayed such undesired deletions. A PCR-based screening of the candidate colonies using test primers flanking the site of deletion is thus necessary.

Using the new method, all deletion experiments were successful, however, insertion efficiency varied widely and was generally lower than that reported by others using λ Red-type recombination. A possible explanation for the generally low number of integrants is the relatively low electroporation efficiency achieved in MG1655. Site to site variations in integration efficiencies are likely to be caused by the particular

nucleotide sequence context of the insertion sites. We note that several of the "difficult" deletions were positioned next to tRNA genes with potential secondary structure.
This method is suitable to construct single or multiple deletions in a wide range of sizes. The largest deletion we constructed covered 82 kb and the smallest possible deletion is about 40 bp, as topological limits are imposed by the duplicated sequences in the latter case.

For the insertion step, pBADαβγ helper plasmid was used in all cases. Phage recombinases responsible for homologous recombination are expressed from the plasmid upon transient induction by arabinose. For expression of I-*Sce*I, two alternatives, based on two different plasmids, were presented. The high-copy-number pKSUC1 expresses I-*Sce*I constitutively and can be used at 37ºC, but curing the cell of it at the end of the process is difficult. On the other hand, pSTKST is easily curable at 37-43 ºC, but the procedure is more time-consuming, because cells must be cultured at 30ºC, and an additional step, derepression of the *tet* promoter is needed to express I-*Sce*I.

Introducing a DSB into the genome induces the SOS response (Walker 1996). This might lead to error-prone DNA synthesis causing incorporation of mutant nucleotides into the chromosome at damaged sites. However, we did not detect (Pósfai et al. 1999, and unpublished results)a significant change in the mutation rate when the method was applied. This might be due to the fact that a single DNA lesion is introduced into the chromosome and there are no other damaged sites serving as potential mutational hotspots.

Repair of a DSB by recombination of duplicated sequences requires a number of repair activities (Kowalczykowski 2000). Limits of primer synthesis dictated the use of short homologies (40-60 bp) as substrates for recombination. RecA is thought to play a minor role in recombination of homologous sequences shorter than 100 bp (Lloyd and Low 1996). However, it was found that presence of RecA was an absolute requirement in this repair process.

5.2 Engineering a reduced *E. coli* genome

Up to now there has not been a concerted study to improve the genome of *E. coli* by large scale reduction. Early work by Squires et. al. used genome deletion of *E. coli* to reduce the number of ribosomal RNA genes for study of their individual functions if any. In *B. subtilis*, Itaya and Tanaka (Itaya and Tanaka 1997) subdivided the genome into two chromosomes.

Using the new method, systematic size reduction of the K-12 MG1655 genome was attempted. Deletion targets were selected primarily by comparing the genomes of MG1655 and O157:H7, and were designed not to disable the cell in its basic metabolic capabilities.

The 4101248 bp genome of MDS31 is smaller than the chromosome of any *E. coli* strain for which data are available (4.4-5.5 Mb) (Bergthorsson and Ochman 1995, Bergthorsson and Ochman 1998). The 31 deletions reduced the genome by removing most of the potentially "selfish" DNA (cryptic prophages, phage remnants, ISs) and a large fraction of genes of tentative or unknown functions. All together, of the 483 ORFs and 2 stable RNA genes deleted, 174 possessed tentative functional classification, and 218 were marked as unknown (Blattner et al. 1997).

Transposons and insertion sequences (ISs) are thought to be the major source of mutations in *E. coli* (Chalmers and Blot 1999). All such mobile genetic elements (44 ISs of MG1655) have been removed from the genome of MDS31, presumably increasing the genetic stability. Experiments measuring the mutation rates are underway to test this hypothesis.

Chromosomal regions identified as K-islands can theoretically carry replacements of essential genes preserved in EDL933 but lost from MG1655. The fact that all 31 regions chosen for deletion could be removed from the MG1655 genome indicates that such non-orthologous gene displacements must be rare events, at least for genes essential under laboratory growth conditions. It may also reflect the high level of paralogy among *E. coli* genes (Riley and Serres 2000) and the robustness of the *E. coli* metabolic network (Edwards and Palsson 2000).

Partial characterization of MDS12 revealed few phenotypic changes, as compared to the parental strain. Growth characteristics and transformability of MDS12 were essentially identical to those of MG1655. The utilization of 379 carbon, nitrogen, phosphorous and sulfur sources was compared by our collaborators in the laboratory of F. R. Blattner, using Biolog Phenotype Microarray plates, and the deletion mutant showed over two-fold reduced growth on 15 compounds. Of the observed phenotypes, the one showing the most marked change can be conclusively attributed to a deleted gene. That is the utilization of glycolate substrate, which is normally metabolized to glyoxylate by the product of the deleted *glcD* gene. Most of the other phenotypes are for compounds whose known catabolic pathways (KEGG database, 2001; http://www.genome.ad.jp/kegg/) are left intact in MDS12. Some of the deleted genes, such as *tynA*, *argF* and *gabT* (whose putative promoter was eliminated) code for products involved in possible alternate or intersecting pathways related to the observed phenotype, but a causal connection is not apparent. In the case of *argF*, the alternative enzyme *argI* is still present in MDS12. A few of the growth defects were observed for compounds that have not been studied in *E. coli* (5-aminovalerate, beta-glycerolphosphate, and parabanic acid). This supports the idea that many unknown ORFs in *E. coli* encode proteins to metabolize uncommon compounds. Similar characterization of MDS31, as well as microarray-based gene-expression analysis of the strains are underway.

Successful size-reduction of the MG1655 genome by 11.6 % underlines the feasibility of a rational, targeted approach in constructing modified bacterial genomes. MDS12 (and MDS31), a strain with a DNA content approaching the “core” *E. coli* genome, appears to retain *E. coli* capabilities important under standard laboratory culturing conditions. Further refinements and deletions could eventually result in a greatly simplified cell displaying fully predictable reactions and programmable functions.

6. Summary

6.1 (English)

Escherichia coli is both a tool and object of study of genome science. As the primary model organism for bacteria it was used to elucidate the biological processes responsible for basic cellular functions. Gene annotations of *E. coli* are regularly transferred to other organisms as their genome sequences are completed. *E. coli* is also used extensively as a genomics tool to propagate DNA subclones for sequencing and for a variety of functional studies in many species. Industrially *E. coli* is used to produce hormones, enzymes and antibiotics. However, in spite of the vast knowledge of the *E. coli* cell, much remains to be learned about the composition of the genome.

Two strains of *E. coli* have been fully sequenced, the K-12 strain MG1655 with 4,638,858 basepairs (Blattner et al. 1997) and two isolates of the enteric pathogen O157:H7 of 5,528,445 basepairs (Hayashi et al. 2001; Perna et al. 2001). K-12 and O157:H7 are phylogenetically distant relatives within the *E. coli* species (Reid et al. 2000, Perna et al. 2001).

Comparison of the genomes of K-12 and O157:H7 revealed a startling pattern in which hundreds of strain-specific "islands" are found inserted into a common "backbone" that is highly conserved (98%) between the two strains. The strain specific islands are termed K-islands or O-islands after the strain from which they were sequenced. The size of the backbone genome is about 3.7 megabasepairs and the total of K-islands amounts to 0.9 Mbp, about 20% of the genome.

Gene loss and horizontal gene transfer were the major genetic processes that shaped the ancestral *E. coli* genome resulting in the spectrum of divergent present-day strains possessing very different arrays of genes (McClelland et al. 2000, Ochman and Jones 2000, Riley and Serres 2000, Perna et al. 2001). The genes contained in the backbone regions generally include basic core functions of *E. coli* that are necessary regardless of environmental niche. Islands contain a disproportionate share of genes that are of unknown function as well as toxins, virulence factors and metabolic capabilities that may be of advantage in the niche to which the strain is adapted. Islands also contain

many transposable elements, phages, cryptic prophages, pseudogenes, gene remnants, and damaged operons. Clearly *E. coli* possesses many dispensable functions that would not be needed in a strain designed for laboratory purposes. Comparative genomics provides the working hypothesis that genes and gene islands that are present in MG1655 and not in other strains may be deleted.

Our laboratory is involved in the development of new methods that allow targeted modifications of the *E. coli* genome. These modifications are then used to answer questions related to gene function and genome structure. Our main goal is to construct a simplified *E. coli* genome by deleting the strain-specific islands not needed for survival under standard laboratory conditions. The cell with a reduced genome can serve as an "improved" model organism and a better biotechnological tool.

This work focused on the development and characterization of two genome-manipulation methods: double-strand break (DSB) stimulated allele replacement and scarless deletion construction. Using these methods, a series of deletions were introduced into the E. coli K-12 genome, and the strain harbouring the reduced genome was partially characterized. The main results were:

- Two variations of the DSB-stimulated method were tested to assess their applicability in generating large deletions, and compared in an experiment aimed at deleting a large K-island of the MG1655 genome. Both variations showed a high efficiency in generating a large deletion (62 kb) in *Escherichia coli* strain MG1655. However, it was found that the method is time-consuming due to the cloning steps.
- A new, rapid and straightforward deletion method was developed. A PCR-generated DNA-fragment was inserted into the genome by λ Red-type homologous recombination, followed by a double strand break-stimulated recombinational repair process, resulting in a scarless deletion. Use of the method required the construction of a set of helper plasmids, including a template plasmid for generating the targeting fragment by PCR, and plasmids for the expression of I-*Sce*I.

- Sensitivity, specificity and other features of the method were tested in several experiments and optimal conditions were established. It was found that (i) 40-60 bp long homologous segments can be used to guide the targeting fragment into the desired locus (ii) both small and large (over 80 kb) deletions can be constructed using the method (iii) the deletion process is recA-dependent.
- Based on comparative genomics, 12 large deletions have been created in MG1655. These deletions included the largest K-islands.
- The deleted strain was partially characterized. Growth rates of MG1655 and the deleted strain, measured both in minimal and rich media, showed no significant difference. Transformation efficiencies of the strains were also similar.
- An additional set of 19 deletions has also been constructed. The total of 31 deletions reduced the genome by removing all major mobile elements (cryptic prophages, phage remnants, 44 ISs) and a large fraction of genes of tentative or unknown functions. All together, of the 483 ORFs and 2 stable RNA genes deleted, 174 possessed tentative functional classification, and 218 were marked as unknown.

MDS31, the strain possessing all 31 deletions, is stripped of all major mobile genetic elements (phages and insertion sequences, primary agents of horizontal gene transfer and genomic rearrangements). This gives us an opportunity to study the role of these mobile elements in the processes of adaptation and evolution.

In general, successful size-reduction of the MG1655 genome by 11.6 % underlines the feasibility of a rational, targeted approach in constructing modified bacterial genomes.

6.2 (Hungarian)

Az *Escherichia coli* baktérium a genomika fontos eszköze, és egyben alanya is. Az alapvető sejtműködési folyamatok felderítésében meghatározó jelentőségű modellszervezetnek bizonyult. Megszekvenált bakteriális genomok analízisekor, a gének azonosítása során összehasonlítási alapként szolgál az *E. coli* genom. A baktérium DNS-klónok gazdasejtjeként fontos szerepet tölt be szekvenálási és génfunkció-vizsgálati munkákban. Hormonok és enzimek termeltetésére az ipar széles körben használja.

Mindezeknek köszönhetően számtalan kísérleti adattal rendelkezünk az *E. coli*-ról, mégis, genomjának összetétele és működése még igen sok ismeretlen részletet rejt.

Két *E. coli* törzs teljes genomszekvenciája ismert. A K-12 MG1655 genom 4 638 858 bp méretű (Blattner et al. 1997), az enterális patogén O157:H7 genomja 5 528 445 bp. Ez utóbbi törzs két izolátumát is megszekvenálták (Hayashi et al. 2001; Perna et al. 2001). A K-12 és az O157:H7 törzs az *E. coli* fajon belül filogenetikailag távol esik egymástól (Reid et al. 2000, Perna et al. 2001) .

A K-12 és az O157:H7 genom öszehasonlítása meglepő eredményt hozott. A két törzsben nagyfokú hasonlóságot mutató (98%-ban azonos) genomgerincet kisebb-nagyobb, törzsspecifikus szigetek százai szakítják meg. Ezeket a tözsspecifikus szigeteket – a törzs nevéből eredően – K-, illetve O-szigeteknek nevezik. A közös genomgerinc mérete mintegy 3,7 Mb. A K-szigetek együttesen 0,9 Mb-t, azaz a K-12 genom mintegy 20%-át teszik ki.

A mai, különféle génkészlettel rendelkező *E. coli* törzsek változatos spektruma két fő genetikai folyamattal alakulhatott ki az ősi *E. coli* genomból: génvesztéssel és horizontális géntranszferrel (McClelland et al. 2000, Ochman and Jones 2000, Riley and Serres 2000, Perna et al. 2001). A közös genomgerinc génjei általában olyan funkciókért felelősek, melyek az alapvető sejtfunkciók biztosításában fontosak, tekintet nélkül az illető törzs által elfoglalt niche-re. A törzsspecifikus szigetek viszont általában toxinokat, virulenciafaktorokat, niche-specifikus metabolikus képességeket kódolnak. Relatíve nagy számban találhatók itt ismeretlen funkciójú gének is, továbbá transzpozonok, fágok, kriptikus profágok, pszeudogének, sérült operonok. Mindezek alapján feltételezhető, hogy az *E. coli* genom számos olyan részletet hordoz, melyre a sejtnek nincsen szüksége laboratóriumi körülmények között. Az összehasonlító genomikai analízis alapján felállítható az a hipotézis, hogy az MG1655 genomból el lehet távolítani a törzsspecifikus géneket és szigeteket, a túléléshez a más törzsekkel közös genomgerinc is elegendő.

Laboratóriumunkban egy sor olyan új módszer került kidolgozásra, melyek segítségével célzott módosításokat lehet végrehajtani az *E. coli* genomban. Ezek a módosítások elsősorban arra irányulnak, hogy egyes gének funkciójáról, illetve a genom struktúrájáról információt nyerjünk. Fő célunk az, hogy a törzsspecifikus szigetek

kiejtésével egy olyan, egyszerűsített *E. coli* genomot készítsünk, mely laboratóriumi körülmények közötti működésében nem sérül. Egy ilyen, redukált genomú sejt "javított" modellszervezetként és biotechnológiai eszközként szolgálhatna.

A disszertáció első részének fókuszában két genommanipulációs módszer – kettős szálú kromoszómatöréssel segített (DSB-stimulált) allélcsere ill. "forradásmentes" deléciókészítés – kipróbálása, fejlesztése és jellemzése áll. A módszerek alkalmazásával aztán egy deléciósorozatot készítettünk az *E. coli* K-12 genomban, majd részlegesen jellemeztük a redukált genomú törzset. Röviden összefoglalva a főbb eredmények a következők:

- Megvizsgáltuk, mennyire alkalmas a DSB-stimulált módszer genomi deléciók létrehozására. A módszer két változatát teszteltük, segítségükkel egy 62 kb méretű deléciót konstruáltunk az MG1655 törzsben. Bár mindkét variáns hatékonynak bizonyult, úgy találtuk, az időigényes klónozási lépések miatt más módszerre van szükség sorozatos deléciók készítéséhez.
- Kidolgoztunk egy új, kellően gyors deléciós módszert. Az eljárás első lépésében egy – PCR segítségével készített – "célzó" DNS-fragmentumot juttatunk a sejtbe, és λ Red típusú homológ rekombinációval beépítjük a genom kiválasztott helyére. A következő lépésben specifikus, kettős szálú kromoszómatörés segítségével rekombinációs javító folyamatokat indukálunk, melyek végső soron létrehozzák a tervezett, "forradásmentes" deléciót. A módszer speciális, a "célzó" DNS-fragmentum létrehozását lehetővé tevő (PCR templát) és az I-*Sce*I meganukleázt expresszáló segédplazmidok konstruálását igényelte.
- Megvizsgáltuk az új módszer érzékenységét, specifitását, és tanulmányoztuk alkalmazásának optimális körülményeit. Megállapítottuk, hogy (i) 40-60 bp hosszúságú homológ szakaszok elegendőek a "célzó" DNS-fragmentum specifikus rekombinációjához (ii) a módszer alkalmas mind kis, mind nagy (akár 82 kb méretű) deléciók készítésére (iii) a deléciót eredményező javító folyamathoz nélkülözhetetlen a RecA fehérje.

- 12 db, összehasonlító genomikai analízissel kiválasztott, nagyméretű deléciót készítettünk az MG1655 törzs genomjában. Ezekkel elsősorban a legnagyobb K-szigeteket ejtettük ki.
- A delécióka hordozó törzset részlegesen jellemeztük. Az eredeti MG1655 és a deléciós törzs növekedési rátája hasonlónak bizonyult mind minimál, mind gazdag táptalajban. Ugyancsak hasonlónak bizonyult a két törzs plazmid DNS transzformációjának hatásfoka.
- A továbbiakban újabb 19 delécióval csökkentettük a deléciós törzs genomjának méretét. Összességében így 31 deléciót hordoz az egyszerűsített genom. Ezek a deléciók eltávolították a genomból az összes jelentősebb mobilis genetikai elemet (kriptikus profágokat, fágmaradványokat, a 44 db inszerciós szekvenciát), továbbá kiejtettek számos ismeretlen funkciójú gént. Mindent összevetve az MG1655 genomból 483 nyitott leolvasási keretet és 2 stabilis RNS-t kódoló szakaszt ejtettünk ki. A deletált gének közül 174-hez feltételezett funkció rendelhető, 218 viszont teljesen ismeretlen.

A 31 delécióval rendelkező törzs (MDS31) mentes az összes olyan genetikai elemtől (fágoktól, inszerciós szekvenciáktól), melyek ismereteink szerint a horizontális géntranszfer események és az intramolekuláris genomi átrendeződések alapvető fontosságú közvetítői. A törzs így egy egyedi eszközt nyújt ahhoz, hogy megvizsgálhassuk a mobilis elemek szerepét az adaptációs és evolúciós folyamatokban.

Az MG1655 genom sikeres, mintegy 11,6 %-os méretredukciója a bakteriális genomok radikális, célzott átalakítását irányzó tervek létjogosultságát demonstrálja.

7. References

- Amundsen S. K., Taylor A. F., Chaudhury A. M. and Smith G. R. recD: the gene for an essential third subunit of exonuclease V. (1986) *Proc. Natl. Acad. Sci. USA* **83**, pp. 5558-5562

- Armstrong K. A., Acosta R., Ledner E., Machida Y., Pancotto M., McCormick M., Ohtsubo H. and Ohtsubo E. A 37 X 10(3) molecular weight plasmid-encoded protein is required for replication and copy number control in the plasmid pSC101 and its temperature-sensitive derivative pHS1. (1984) *J. Mol. Biol.*, **175**, pp. 331-347

- Bachellier S., Gilson E., Hofnung M., Hill C. W. in Escherichia coli *and* Salmonella *Cellular and Molecular Biology*, F. C. Neidhardt *et al.*, Eds. (ASM Press, Washington, DC, 1996) vol. 2, pp. 2012-2040

- Bergthorsson, U. and Ochman, H. Heterogeneity of genome sizes among natural isolates of *Escherichia coli.* (1995) *J. Bacteriol.* **177:** 5784-5789

- Bergthorsson, U. and Ochman, H. Distribution of chromosome length variation in natural isolates of *Escherichia coli.* (1998) *Mol. Biol. Evol.* **15:** 6-16

- Biek D. P. and Cohen S. N. Identification and characterization of recD, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. (1986) *J. Bacteriol.* **167**, pp. 594-603

- Blattner F. R., Plunkett G., III, Bloch C. A., Perna N. T., Burland V., Riley M., Collado-Vides J., Glasner J. D., Rode C. K., Mayhew G. F et al. The complete genome sequence of *Escherichia coli* K-12. (1997) *Science* **277:** pp. 1453-1474

- Blomfield I. C., Vaughn V., Rest R. F. and Eisenstein B. I. Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. (1991) *Mol. Microbiol.*, **5:** pp. 1447-1457.

- Campbell A. M. in Escherichia coli and Salmonella Cellular and Molecular Biology, F. C. Neidhardt *et al.*, Eds. (ASM Press, Washington, DC, 1996) vol. 2, pp. 2041-2046

- Chalmers, R. and Blot, M. Insertion sequences and transposons. (1999) In *Organization of the prokaryotic genome* (ed. R.L. Charlebois), pp. 151-169. ASM Press, Washington, D.C.

- Clark A. J. and Margulies A. D. Isolation and characterization of recombination deficient mutants of *Escherichia coli* K12. *Proc. Natl. Acad. Sci. USA* **53:** pp. 451-459
- Clark A. J. and Sandler S. J. Homologous genetic recombination: the pieces begin to fall into place. (1994) *Crit. Rev. Microbiol.* **20**, pp. 125-142

- Cosloy S. D. and Oishi M. Genetic transformation in *Escherichia coli* K12. (1973) *Proc. Natl. Acad. Sci. USA* **70**, pp. 84-87

- Datsenko K. A., Wanner B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. (2000) *Proc Natl Acad Sci U S A.* Jun 6; **97**(12)**:** pp. 6640-6645

- Dean D. A plasmid cloning vector for the direct selection of strains carrying recombinant plasmids. (1981) *Gene*, **15**, pp. 99-102

- Edwards, J.S. and Palsson, B.O. Robustness analysis of the *Escherichia coli* metabolic network. (2000) *Biotechnol. Prog.* **16:** pp. 927-939

- Filutowicz M., Davis G., Greener A. & Helinski D. Autorepressor properties of the pi-initiation protein encoded by plasmid R6K. (1985) *Nucleic Acids Res.* **13**, pp. 103-114

- Filutowicz M., Dellis S., Levchenko I., Urh M., Wu F. & York D. Regulation of replication of an iteron-containing DNA molecule. (1994) in *Progress in Nucleic Acid Research and Molecular Biology*, eds. Cohn, W. & Moldave, K. (Academic, San Diego), Vol. 48, pp. 239-273

- Filutowicz M., McEachern M. J. & Helinski D. R. Positive and negative roles of an initiator protein at an origin of replication. (1986) *Proc. Natl. Acad. Sci. USA* **83**, pp. 9645-9649

- Filutowicz M., Uhlenhopp E. & Helinski D. R. Binding of purified wild-type and mutant-pi initiation proteins to a replication origin region of plasmid R6K. (1985) *J. Mol. Biol.* **187**, pp. 225-239

- Frey J., Chandler M., Caro L. The effects of an *Escherichia coli* dnaAts mutation on the replication of the plasmids colE1 pSC101, R100.1 and RTF-TC. (1979) *Mol Gen Genet.* Jul 13;**174**(2)**:** pp. 117-26

- Gay P., LeCoq D., Steinmetz M., Berkelman T. and Kado C. I. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. (1985) *J. Bacteriol*., **164**, pp. 918-921

- Germino J. & Bastia D. Interaction of the plasmid R6K-encoded replication initiator protein with its binding sites on DNA. (1983) *Cell* **34**, pp. 125-134

- Germino J. & Bastia D. The replication initiator protein of plasmid R6K tagged with beta-galactosidase shows sequence-specific DNA-binding. (1983) *Cell* **32**, pp. 131-140

- Hamilton C. M., Aldea M., Washburn B. K., Babitzke P. and Kushner S. R. New method for generating deletions and gene replacements in *Escherichia coli*. (1989) *J. Bacteriol*., **171:** pp. 4617-4622

- Hasunuma K, Sekiguchi M. Effect of dna mutations on the replication of plasmid pSC101 in *Escherichia coli K-12*. (1979) *J Bacteriol.* Mar; **137**(3)**:** pp. 1095-9

- Hasunuma K., Sekiguchi M. Replication of plasmid pSC101 in *Escherichia coli* K12: requirement for dnaA function. (1977) *Mol Gen Genet.* Sep 9; **154**(3)**:** pp. 225-30

- Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C. G., Ohtsubo E., Nakayama K., Murata T et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. (2001) *DNA Res.* **8:** pp. 11-22

- Howard-Flanders P. and Theriot L. Mutants of Escherichia coli K-12 defective in DNA repair and in genetic recombination. (1966) *Genetics*. Jun; **53**(6)**:** pp. 1137-50

- Itaya, M. and Tanaka, T. Experimental surgery to create subgenomes of *Bacillus subtilis* 168. (1997) *Proc. Natl. Acad. Sci.* **94:** 5378-5382

- Jasin M., and Schimmel P. Deletions of an essential gene in *Escherichia coli* by site-specific recombination with linear DNA fragments. (1984) *J. Bacteriol.* **159:** pp. 783-786

- Kelley W. L. & Bastia D. Replication initiator protein of plasmid R6K autoregulates its own synthesis at the transcriptional step. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2574-2578

- Kitamura, K., Torii, Y., Matsuoka, C., and Yamamoto, K. DNA sequence changes in mutations in the *tonB* gene on the chromosome of *Escherichia coli* K12: Insertion elements dominate the spontaneous spectra. (1995) *Jpn. J. Genet.* **70:** 35-46

- Kowalczykowski S.C. Initiation of genetic recombination and recombination-dependent replication. (2000) *Trends Biochem. Sci.* **25:** pp. 156-165

- Link A. J., Phillips D. and Church G. M. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. (1997) *J. Bacteriol*., **179:** pp. 6628-6237

- Lloyd R. G. and Low B. L. (1996) In Neidhardt,F.C. (ed.), *Escherichia coli and Salmonella*.ASM Press, Washington, DC, Vol. II, pp. 2237

- Lorenz M. G. and Wackernagel W. Bacterial gene transfer by natural genetic transformation in the environment. (1994) *Microbiol. Rev.* **58**, pp. 563-602

- Maloy S. R., and Nunn W. D. Selection for loss of tetracycline resistance by *Escherichia coli*.(1981) *J. Bacteriol.* **145:** pp. 1110–1112

- Maloy S. R., Stewart V. J., and Taylor R. K. Genetic analysis of pathogenic bacteria. (1996) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

- Marinus M. G., Carraway M., Frey A. Z., Brown L. and Arraj J. A. Insertion mutations in the *dam* gene of *Escherichia coli* K-12.(1983) *Mol. Gen. Genet.* **191:** pp. 288-289

- McClelland, M., Florea, L., Sanderson, K., Clifton, S.W., Parkhill, J., Churcher, C., Dougan, G., Wilson, R.K., and Miller, W. Comparison of the *Escherichia coli* K-12 genome with sampled genomes of a *Klebsiella pneumoniae* and three *Salmonella enterica* serovars, Typhimurium, Typhi and Paratyphi. (2000) *Nucleic Acids Res.* **28:** 4974-4986

- Miller J. H. A short course in bacterial genetics: A laboratory manual and handbook for Escherichia coli and related bacteria., (1992) pp. 271-273. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

- Miller V. L. and Mekalanos J. J. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*.(1988) *J. Bacteriol.* **170:** pp. 2575–2583

- Monteilhet C., Perrin A., Thierry A., Colleaux L. and Dujon B. Purification and characterization of the in vitro activity of I-*Sce*I, a novel and highly specific endonuclease encoded by a group I intron. (1990) *Nucleic Acids Res.* **18:** pp. 1407-1413

- Murphy K. C. Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. (1998) *J. Bacteriol.* **180**, 2063-2071

- Muyrers J. P., Zhang Y., Testa G. and Stewart A. F. Rapid modification of bacterial artificial chromosomes by ET- recombination.(1999) *Nucleic Acids Res.* **27:** pp. 1555-1557

- Ochman H. and Jones I. B. Evolutionary dynamics of full genome content in *Escherichia coli.*(2000) *EMBO J.* **19:** pp. 6637-6643

- O'Connor M., Peifer M., Bender W. Construction of large DNA segments in *Escherichia coli*. (1989) *Science*. Jun 16; **244**(4910)**:** pp. 1307-12

- Perna N. T., Plunkett G., III, Burland V., Mau B., Glasner J. D., Rose D. J., Mayhew G. F., Evans P. S., Gregor J., Kirkpatrick H. A et al. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. (2001) *Nature* **409:** pp. 529-533

- Pósfai G., Kolisnychenko V., Bereczki Z., Blattner F. R. Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome.(1999) *Nucleic Acids Res.* Nov 15; **27**(22): pp. 4409-15

- Pósfai G., Koob M. D., Kirkpatrick H. E. and Blattner F. R. Versatile insertion plasmids for targeted genome manipulations in bacteria: isolation, deletion, and rescue of the pathogenicity island LEE of the *Escherichia coli* O157:H7 genome. (1997) *J. Bacteriol*., **179**, pp. 4426-4428

- Poustka A., Rackwitz H. R., Frischauf A. M., Hohn B., Lehrach H. Selective isolation of cosmid clones by homologous recombination in *Escherichia coli.*(1984) *Proc Natl Acad Sci U S A.* Jul; **81**(13)**:** pp. 4129-33

- Reid S.D., Herbelin C. J., Bumbaugh A. C., Selander R. K., and Whittam T. S. Parallel evolution of virulence in pathogenic *Escherichia coli*. (2000) *Nature* **406:** 64-67

- Retallack D. M., Johnson L. L., Friedman D. I. (1994) *J. Bacteriol.* **176**, 2082 Kirby J. E., Trempy S. Gottesman, *ibid.*, p. 2068

- Richmond, C.S., Glasner, J.D., Mau, R., Jin, H., and Blattner, F.R. Genome-wide expression profiling in *Escherichia coli* K-12. (1999) *Nucleic Acids Res*. **27:** 3821-3835

- Ried J. L. and Collmer A. An *nptI-sacB-sacR* cartridge for constructing directed, unmarked mutations in gram-negative bacteria by marker exchange-eviction mutagenesis.(1987) *Gene* **57:** pp. 239–246

- Riley, M. and Serres, M.H. Interim report on genomics of *Escherichia coli.*(2000) *Annu. Rev. Microbiol.* **54:** 341-411

- Russell C. B. and Dahlquist F. W. Exchange of chromosomal and plasmid alleles in Escherichia coli by selection for loss of a dominant antibiotic sensitivity marker. (1989) *J. Bacteriol*., **171**, pp. 2614-2618

- Sambrook J., Fritsch E.F. and Maniatis T. Molecular cloning: A laboratory manual. (1989) pp. 1.82-1.84., A.1-A.2 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

- Schaechter, M. and Neidhardt, F.C. (1987) Introduction. In *Escherichia coli and Salmonella* (ed. F.C. Neidhart et al.), pp. 1-2. ASM Press, Washington, D.C.

- Skorupski K., and Taylor R. K. Positive selection vectors for allelic exchange. (1996) *Gene* **169:** pp. 47–52

- Snyder L., and Champness W. Molecular genetics of bacteria.(1997) ASM Press, Washington, D.C.

- Stalker D. M., Kolter R. and Helinski D.R. Plasmid R6K DNA replication. I. Complete nucleotide sequence of an autonomously replicating segment. (1982) *J. Mol. Biol.*, **161**, pp. 33-43

- Stibitz S., Black W. and Falkow S. The construction of a cloning vector designed for gene replacement in *Bordetella pertussis*. (1986) *Gene* **50:** pp. 133–140

- Walker, G.C. (1996) The SOS response of *Escherichia coli*. In *Escherichia coli and Salmonella* (ed. F.C. Neidhart et al.), pp. 1400-1416. ASM Press, Washington, D.C.

- Winans S. C., Elledge S. J., Krueger J. H. and Walker G. C. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*.(1985) *J. Bacteriol.* **161:** pp. 1219-1221

- Woodcock D. M., Crowther P. J., Doherty J., Jefferson S., DeCruz E., Noyer-Weidner M., Smith S. S., Michael M. Z. and Graham M. W. Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. (1989) *Nucleic Acids Res.*, **17**, pp. 3469-3478

- Yu D., Ellis H. M., Lee E. C., Jenkins N. A., Copeland N. G. and Court D. L. An efficient recombination system for chromosome engineering in *Escherichia coli.* (2000) *Proc. Natl. Acad. Sci.* **97:** pp. 5978-5983

- Zhang Y., Buchholz F., Muyrers J. P. and Stewart A. F. A new logic for DNA engineering using recombination in *Escherichia coli.* (1998) *Nat. Genet.* **20:** pp. 123-128

7. List of author's publications

1. Kolisnychenko, V., Plunkett, G. 3rd, Herring, C.D., Fehér, T., Pósfai, J., Blattner, F.R., Pósfai, G. "Engineering a reduced *Escherichia coli*"
   Genome Research. April 2002; 12(4); 640-647

2. Pósfai, G., Kolisnychenko, V., Bereczki, Z., Blattner, F.R. "Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome".
   Nucleic Acids Research. November 1999; 27(22): 4409-15

3. Poster presentation "Az *E. coli* genom egyszerűsítése" on the VI Hungarian Biochemical Society Annual Meeting, 14-17 May 2002, Sárospatak, Hungary

4. Poster presentation "Engineering a reduced *Escherichia coli*" on the 168th AAAS Annual Meeting, 14-19 February 2002, Boston, MA, USA

5. Poster presentation "Az *E. coli* genom hasítása: a javító mechanizmusok felhasználása tetszőleges gén-manipulációkhoz" on the IV Hungarian Biochemical Society Annual Meeting, 10-13 May 1999, Eger, Hungary